Cloning and characterisation of pullulanase type 1 from thermopallium natronophilum

Koo, Mei Keng

Award date:
2004

Awarding institution:
University of Bath

Cloning and Characterisation of Pullulanase Type I from *Thermopallium natronophilum*

Submitted by Mei Keng Koo

for the degree of Ph.D

of the University of Bath

2004

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

谨以此文献给我的爸爸妈妈：

To Mama and Papa

And in memory of my grandpas

Science is a wonderful thing if one does not have to earn one's living at it.

~Albert Einstein~

1879-1955

Acknowledgements

First and foremost I would like to thank my parents and my siblings (Be How and Be Theng) for showering me with love, believing in me and giving me the moral support that I needed. Following this my grandmothers, aunts and uncles, cousins, Peggy, family friends for their love, kindness and support all these years. My guardians in England, Uncle Norman and Aunt.Chris for all their encouragement. (those Easter and Christmas chocolates sure cheered me up!)

Secondly, I would like to thank both my supervisors, Mike and David for giving me the opportunity to embark on a PhD. degree, for believing in me and helping me through my struggle; and for the patience and kindness that they have shown me. I would also like to thank my internal supervisor, James for being so supportive and understanding. Not forgetting the CER staff, Alan (who is a God sent), Alex, Susan, Liz and Richard for your help, guidance and friendship.

I would also like to say thank you to my colleagues (past and present), Carlos, Ursula, Queenie (for kicking my A*** all the time), Debbie, Becky, Tree, Tina, Nathelie Da Crazy French Gal, Dina (the sister that I've never had), Harry and Caroline for spicing up lab 1.33 (Oh, how I miss the Mai-Thai days!).

All my thanks and love to my mates in Bath and London, Jean, Evelyn, Sigrid, Carol, Myriam, Khue, Ling, Rowena, Caroline, Bridget, Fazil, Belinda, Dizery, Jing, Sabirah, Mamdoh, Michael, Harumi, Mario, Jane, Yahui, Kenny, Huan-Lin, John, Kartyk, everybody from lab 1.28 (past and present) and all those that I did not have the chance to mention, I thank you. A special thank you to my mates in all things devious Zen, Hong and Shalin for giving me the best escapades in year 2002.

Also, not forgetting all the staff from Oxfam, especially Max, Vince, Jan, Mary, Caroline, Dita, Dave, James, Peggy for making my Saturday mornings so enjoyable.

A very special thank you to Zhen Yu for your love and support for the past two years. Thanks for helping me finish this race.

Last but not the least, University of Bath for funding me.

Abstract

Thermopallium natronophilum was first isolated from a soda lake, Lake Bogoria in the Kenyan Tanzanian Rift in 1990. One of the starch hydrolysing enzymes it produces is known as pullulanase. The native enzyme has been purified to homogeneity using gel filtration, anion exchange chromatography and affinity chromatography. The enzyme has been identified to be a type I pullulanase which specifically hydrolyses α-1,6 glycosidic linkages (Thompson, 1998).

The mature type I pullulanase is a monomeric protein with an estimated Mr of 87,000Da. The enzyme is found to have an optimum temperature of 90°C and an optimum pH of 7.5. Due to the ability of this enzyme to remain active at high temperature and high pH, it can be exploited in several industries, e.g. detergent industry, textile industry and bakery industry.

From determined internal amino acid sequence of the native protein, and consensus sequences of pullulanases, specific and non-specific oligonucleotide primers were designed and used to PCR amplify the majority of the *T. natronophilum* gene sequence. The N and C terminal sequences were obtained via nested PCR that involved the use of random and specific primers, in each case one of which was biotinylated. Selection of the desired amplified fragments was carried out by immobilisation of the PCR products on streptavidin-coated paramagnetic beads, and from these the N and C terminal sequences of the gene were obtained.

Following this, the pullulanase gene without the signal peptide was cloned into an expression vector, pET28a. The recombinant pullulanase was successfully expressed and purified to homogeneity with affinity and anion exchange chromatography. The recombinant pullulanase was then characterised. The pH optimum and substrate specificity were found to be in agreement with the native pullulanase. Unfortunately, the recombinant pullulanase is less thermal stable than the native pullulanase.

Abbreviations

α-cyclodextrin	cyclohexaamylose
Amp	ampicillin
ANSA	3-amino, 5nitrosalicyclic acid
bp	base pairs
BSA	bovine serum albumin
CTAB	hexadecyltrimethyl ammonium bromide
CGTase	cyclomaltodextrin-glucanotransferase
Da	Daltons
dNTP	di-deoxy-nucleotide triphosphate
dATP	di-deoxy-adenosine triphosphate
dCTP	di-deoxy-cytosine triphosphate
dGTP	di-deoxy-guanosine triphosphate
dTTP	di-deoxy-thymidine triphosphate
DIG	digoxigenin
DNA	deoxyribosenucleic acid
DNSA	3,5-dinitrosalicyclic acid
dNTP	deoxy-nucleotide triphosphate
EDTA	(disodium) ethylenediamine tetraacetate
FP	flanking primer
FPLC	fast protein liquid chromatography
gDNA	genomic DNA
HEPES	(N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulphonic acid])
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilobase pairs
kDa	kilo Daltons
LB	Luria-Bertani
MES	(2-[N-morpholino]ethanesulphonic acid)
MOPS	3-(N-morpholino) propanesulphnic acid
M_r	relative molecular weight
MQ H_2O	Milli Q water
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
RNAse A	ribonuclease A (EC 3.1.27.5)
rRNA	ribosomal ribonucleic acid
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulphate
SSC (20×)	3M NaCl, 0.3M sodium citrate (pH7.0)
TAE	40mM Tris-acetate, 1mM EDTA
TE	10mM Tris-HCl, 1mM EDTA
TEMED	N, N', N', N' tetramethylethylene diamine
TIM	triosephosphate isomerase
Tris	Tris-(hydroxymethyl)-,ethylamine
UV	ultraviolet

X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
v/v	volume per volume ratio
w/v	weight per volume ratio

Amino Acid	3-Letter Code	1-Letter Code	Hydrophobicity
Alanine	Ala	A	0.616
Cysteine	Cys	C	0.680
Aspartate	Asp	D	0.028
Glutamate	Glu	E	0.043
Phenylalanine	Phe	F	1.00
Glycine	Gly	G	0.501
Histidine	His	H	0.165
Isoleucine	Ile	I	0.943
Lysine	Lys	K	0.283
Leucine	Leu	L	0.943
Methionine	Met	M	0.738
Asparagines	Asn	N	0.236
Proline	Pro	P	0.711
Glutamine	Gln	Q	0.251
Arginine	Arg	R	0.000
Serine	Ser	S	0.359
Theronine	The	T	0.450
Valine	Val	V	0.825
Tryptophan	Trp	W	0.878
Tyrosine	Tyr	Y	0.880

Single Letter Code Used in Degenerate Primers

Single Letter Code	Nucleotides
M	A, C
R	A, G
W	A, T
S	G, C
Y	C, T
K	G, T
V	A, G, C
H	A, C, T
D	A, G, T
B	G, C, T
N	A, G, C, T

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Overview

The initial aim of this project was to identify the gene sequence of a type I pullulanase from the thermoalkalophile *Thermopallium natronophilum*. Following this, the cloning and expression of the pullulanase gene were carried out. Preliminary results suggested that the recombinant enzyme was active and thus this led to the purification of the recombinant protein. Consequently, the characterisation of the recombinant pullulanase was carried out and compared to the native enzyme in the aspects of optimum temperature, optimum pH, thermal stability and substrate specificity.

In this chapter, the thermoalkalophile, *Thermopallium natronophilum*, will first be introduced. Aspects touched on will be the origin, growth conditions, morphology and phylogeny of this bacterium. The growth of the bacterium experimentally will be outlined in Chapter 3.

Following this, the enzyme, type I pullulanase is discussed. This enzyme is a glycosyl hydrolase that belongs to the glycoside hydrolase family 13, also known as the α-amylase family. The characteristics of the glycoside hydrolase family 13 will also be examined. Subsequently, the substrate and the enzymatic hydrolysis will be presented.

Finally, the potential biotechnological application of this enzyme in various industries will be outlined, especially the detergent industry, textile industry and baking industry.

1.2 The Microorganism

The microorganism from which the native pullulanase was originally isolated is *Thermopallium natronophilum*. This bacterium was first isolated in 1992 from a soda lake, Lake Bogoria, located in the Kenyan-Tanzanian Rift (Duckworth *et al*, 1996; Wiegel, 1998; Jones *et al.*, 1998). *Thermopallium natronophilum* has a growth temperature range of 52°C to 78°C and a pH range of 7.2 to 10.5 (Thompson, 1998). This microbe is found to have an optimum growth temperature of 70°C and pH optimum of 9.5. It is an obligate anaerobe and stained Gram negative (Jones *et al*, 1998).

Morphologically, this bacterium resembles a *thermotogale* known as *Thermotoga maritima*. This is a rod-shaped cell surrounded by a porous proteinaceous coat known as the toga (Fig.1.1, Fig.1.2). The toga is like a sheath that attaches firmly to the cylindrical body of the bacterium but inflates up at both the poles of the cell (Rachel *et al.*, 1990). The toga wall contains a major component, a 42 kDa protein that organises itself into a trimer. This trimeric protein was shown to resemble the OmpF trimers found in *Escherichia coli*.

Even though preliminary results indicated that *T.natronophilum* is related to *T.maritima*, according to 16S rRNA sequence similarity test, it is more closely related to another *thermotogale* known as *Fervidobacterium* (Fig.1.3) (Duckworth *et al.*, 1996). In both *T.maritima* and *Fervidobacterium pennavorans Ven*5, the type I pullulanase were identified. Due to the close family lineage, enzymes from these two strains have been used as models of comparison.

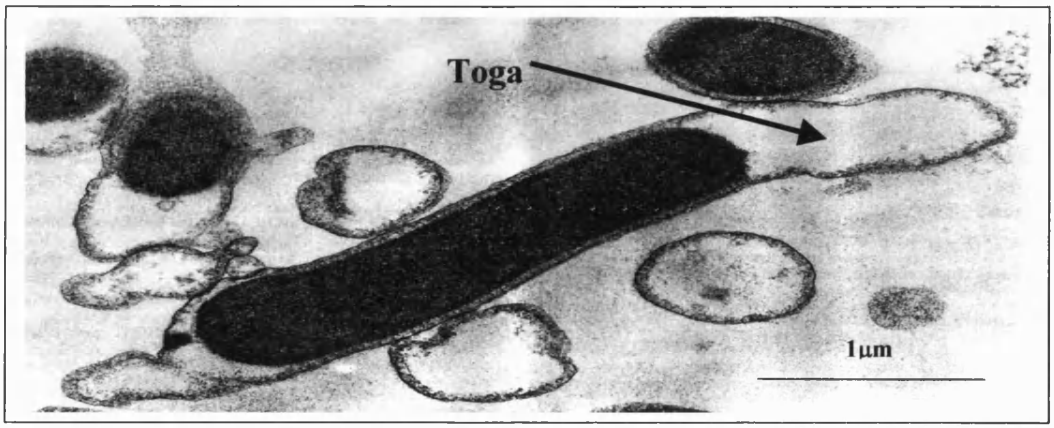


Fig.1.1: Thin section of *Thermotoga maritima* (Huber *et al.*, 1986).

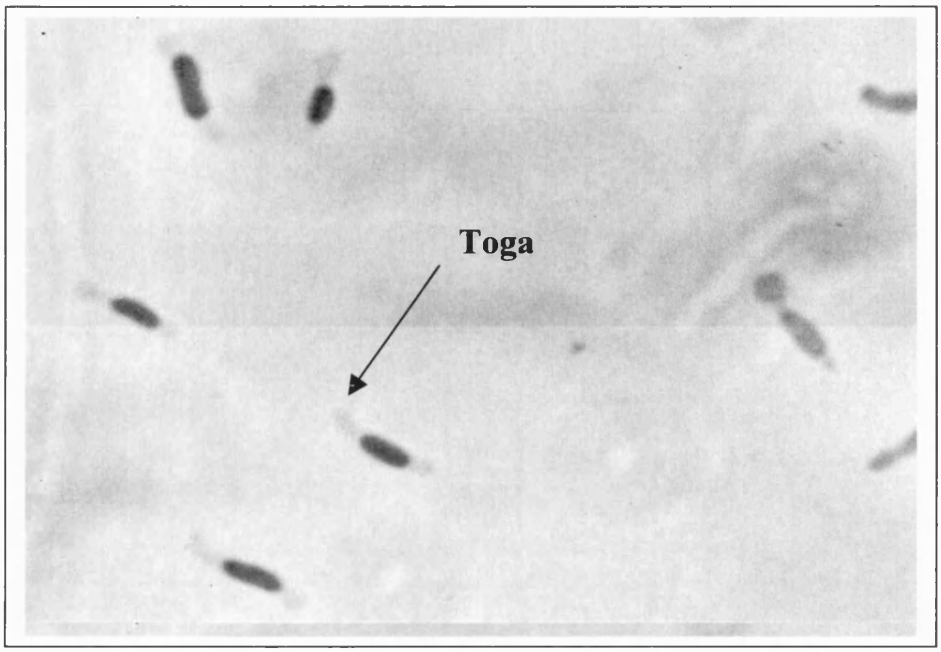


Fig.1.2: *Thermopallium natronophilum* under light microscopy 1000× magnification.

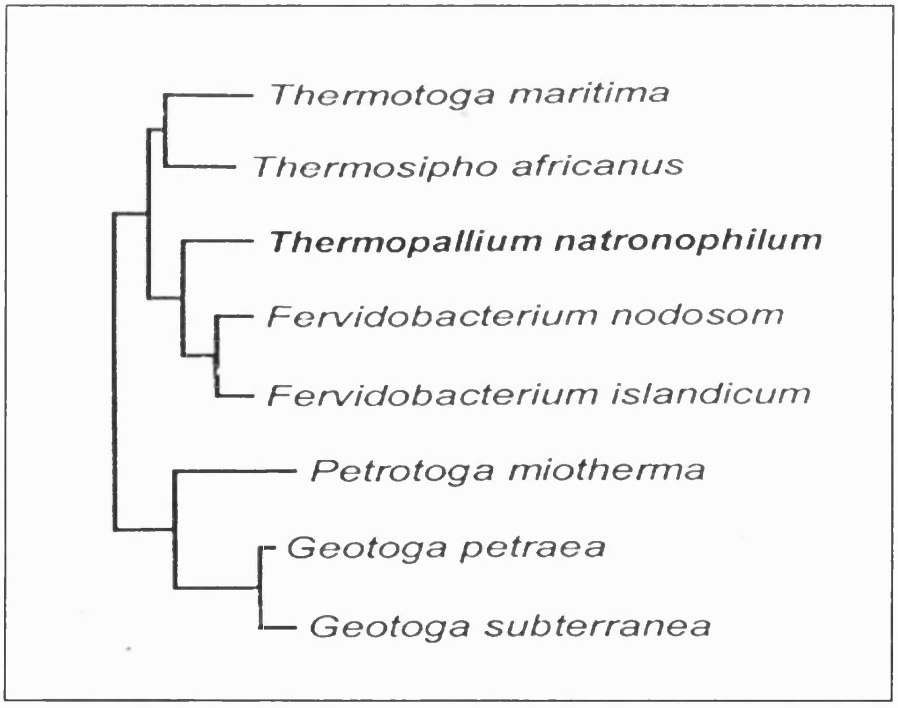


Fig.1.3: Phylogenetic tree: According to 16S rRNA analysis, *Thermopallium natronophilum* is closer to *Fervidobacterium* than to *Thermotoga maritima* (Duckworth *et al.*, 1996)

1.3 The Enzyme

1.3.1 The Glycoside Hydrolase Family 13

Pullulanase type I is categorised under the glycosyl hydrolase family 13 (Carbohydrate Active Enzymes Database, CAZy website). The glycoside hydrolase family 13 is also known as the α-amylase family. Members of the α-amylase family generally have structural or functional similarities. Members of this family adopt the infamous TIM barrel, where eight helices surround a barrel of eight parallel β-strands (Fig.1.4). The enzymes from glycoside hydrolase family 13 catalyse either the hydrolysis or the transglycosylation of α-linked glucans without disrupting the anomeric configuration (MacGregor *et al.*, 2001). A schematic diagram that encompasses the members of this family with their type of catalytic reactions can be found in Fig.1.5. A full list of the different enzymes categorised under this family is listed in Table 1.1. Of all the different enzymes in the glycosyl family 13, α-amylase (EC 3.2.1.1) is the most widely studied member.

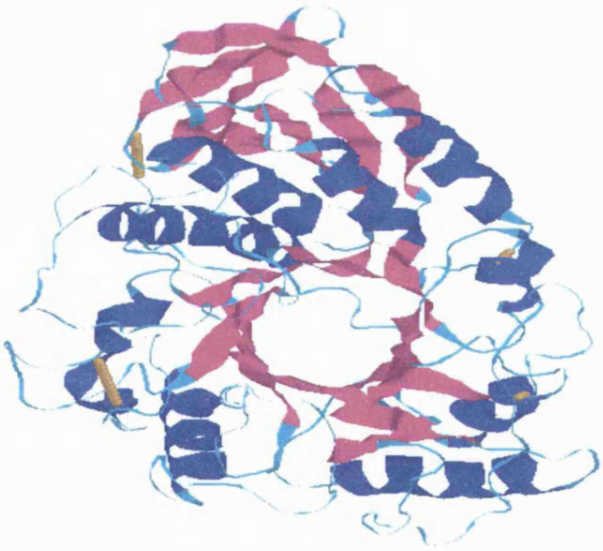


Fig.1.4: α-Amylase from *Aspergillus oryzae*, PDB ID: 2TAA (Matsuura *et al.*, 1984). The eight beta strands forming the cylindrical barrel have been painted violet and the eight alpha-helices surrounding the barrel have been painted blue. Disulphide bridges have been coloured yellow ochre. This is done using Protein Explorer 2.

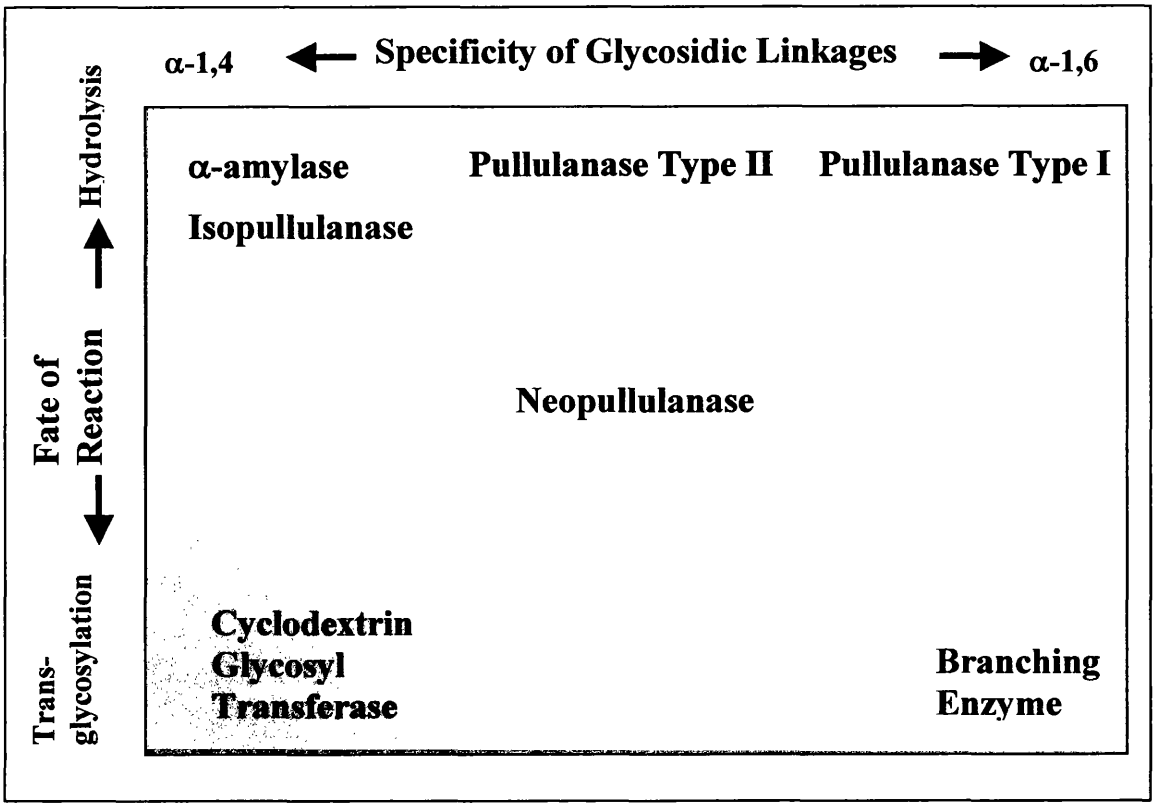


Fig.1.5: This schematic diagram illustrates the types of reaction that define the enzymes found in the α-amylase family (glycosyl hydrolase family 13). α-Amylase (EC 3.2.1.1) hydrolyses starch at the α-1,4 glycosidic linkages while pullulanase type I (EC 3.2.1.41) hydrolyses starch at α-1,6 glycosidic linkages. Cyclodextrin glycosyl transferase (EC 2.4.1.19) catalyses the formation of α-1,4 linkages while branching enzyme, 1,4-α-D-Glucanano transferase, catalyses the formation of α-1,6 linkages via trans-glycosylation. Adapted from www.glycoform.gr.jp/science/word/saccharide/SA-BO5E.html. (Saha *et al.*, 1991)

Table 1.1: The list of members with known activities in the glycoside hydrolase family 13 Adapted from CAZy (Carbohydrate Active Enzymes Database) website.

EC Number	Enzymes
EC 3.2.1.1	α-amylase
EC 3.2.1.41	Pullulanase
EC 2.4.1.19	Cyclomaltodextrin glucanotransferase
EC 3.2.1.54	Cyclomaltodextrinase
EC 3.2.1.93	Trehalose-6-phosphate hydrolase
EC 3.2.1.10	Oligo-α-glucosidase
EC 3.2.1.133	Maltogenic amylase
EC 3.2.1.135	Neopullulanase
EC 3.2.1.20	α-glucosidase
EC 3.2.1.60	Maltotetraose-forming α-amylase
EC 3.2.1.68	Isoamylase
EC 3.2.1.70	Glucodextranase
EC 3.2.1.98	Maltohexaose-forming α-amylase
EC 2.4.1.18	Branching enzyme
EC 5.4.99.16	Trehalose synthase
EC 2.4.1.25	4-α-glucanotransferase
EC 3.2.1.	Maltopentaose-forming α-amylase
EC 2.4.1.4	Amylosucrase
EC 2.4.1.7	Sucrose phosphorylase

1.3.2 Glycoside Hydrolase Family 13: The Multi Domain Protein

Structurally, members in the glycoside hydrolase family 13 contain three main domains, with domain A being the TIM barrel structure, which makes up the core of the enzyme. This barrel is composed of a minimum of 200 residues and in an ideal situation it is composed of eight parallel β-sheets forming the barrel and surrounding the barrel are eight α-helices. All the eight β strands are situated adjacent to each other and hydrogen bonds can be established between them. The core of the barrel is hydrophobic (Fersht, 1999; Machius *et al.*, 1995).

This is followed by domain B, which is a large loop extending from the third β-strand to the third α-helix. Domain B is β-rich in structure and varies considerably in size and structure in all amylases. The last domain, C, is composed entirely of β-strands that sandwich together to give a Greek key motif (Fig.1.6) (Nielsen *et al.*, 2000). This domain is believed to stabilise the catalytic domain A by shielding the hydrophobic amino acids from surrounding solvents. It was also proposed that this domain might also assist the enzyme in substrate binding.

Three further domains have been categorised and given the names domain D, E and N. Not all three domains are reported in all members in the glycoside hydrolase family 13. Properties of domain D remain elusive but domain E has been proposed to play a vital role in binding granular starch. It has been proposed that the properties of two latter domains are very much influenced by domain A and B (Veen *et al.*, 2000). Domain N has been found in neopullulanase of *Bacillus stearothermophilus* (Hondoh *et al.*, 2003) and α-amylase II of *Thermoactinomyces vulgaris* R-47 (Kamitori *et al.*, 1999). In both enzymes, domain N is found to be involved in dimer formation.

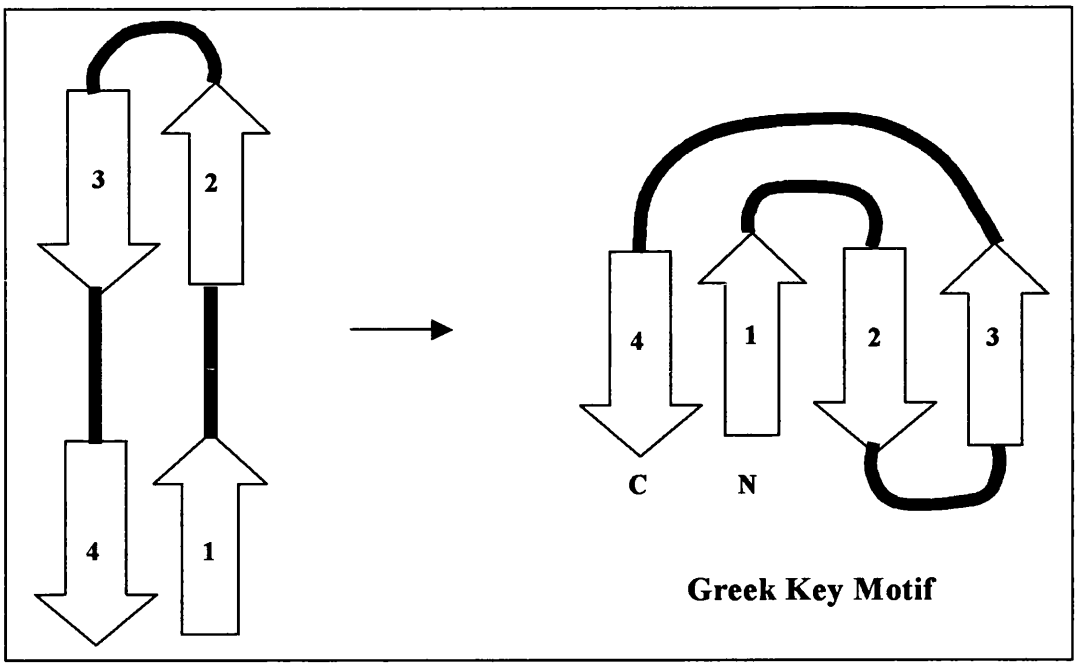


Fig. 1.6: Greek key motif of domain C adapted from Fersht (1999).

1.3.3 Identifying Homologous Regions and Catalytic Residues in Glycoside Hydrolase Family 13

Enzymes belonging to the α-amylase family contain four highly conserved regions (Janse *et al.*, 1993; Jesperson *et al.*, 1993, MacGregor .*et al.*, 2001). In Table 1.2, several enzymes from different strains have been used for the identification of these four conserved regions. Enzymes used for alignment are pullulanase I from *T.natronophilum*, *F.pennavorans* Ven5, *Thermus sp.*, *Klebsiella aerogenes*, *T.maritima;* α-amylase from *T.maritima*, *Bacillus megaterium*, *Dictyoglomus thermophilum*; neopullulanase from *Bacillus stearothermophilus* and *Thermoactinomyces vulgaris.*

All enzymes in the α-amylase family are thought to have a similar mechanism of action. In TAKA amylase (α-amylase from *Aspergillus oryzae*) three carboxylic acids (i.e. two aspartic acids and one glutamic acid) have been identified as important for catalysis. These three carboxylic acids are found to be highly conserved throughout

the α-amylase family. The first aspartic acid has been identified as the second aspartic acid in Region II and the second aspartic acid is located in Region IV. The glutamic acid is found to be in Region III. The asterisk annotating these three carboxylic acids has been highlighted in red in Table 1.2. The function of these carboxylic acids will be further elaborated in the mode of enzymic hydrolysis in a later section.

The first aspartic acid (D) in Region I has been proposed to form a salt bridge with arginine (R), the fourth conserved residue in Region II, to ensure a fully functional active site in cyclodextrin glycosyltransferase, CGTase (Klein *et al.*, 1992). It has been assumed that this is also the case for α-amylase. Vihinen *et al.* (1990) carried out mutational studies on the arginine in α-amylase of *B.stearothermophilus* and a loss of 88% enzymatic activity was observed.

The second tyrosine (Y) located in Region V, present in all enzymes catalysing α-1,6 hydrolysis and in some amylases and CGTases, has been proposed to form a hydrogen bond with arginine in Region II (Vihinen *et al.*, 1990) or with the aspartic acid in region I through a water molecule in CGTases (Klein *et al.*, 1992).

The highly conserved valine residue in region I has been postulated to be involved in enzyme substrate interactions (Jesperson, 1993). The two highly conserved histidines in Region I and IV have been proposed to bind to the non-reducing end of the glucosyl residue (MacGregor *et al.*, 1988; Nakamura *et al.*, 1993). Kuriki *et al.* (1991) have shown that mutation of either of these histidines to a glutamine suppressed the activity of α-1,6 hydrolysis.

Table 1.2: The four highly conserved regions normally found in members from glycosyl family 13. Region V is highly conserved in pullulanase type I. The * shows the highly conserved amino acids across the different enzymes from the different strains. The * indicates the conserved residues important for catalytic activity.

Enzyme	Organism	Region I	Region II	Region III	Region IV	Region V	References
Pullulanase I	*F.pennavorans* Ven5	DMVFPH	DGFRFDQMGL	EPWG	YVEVHD	YNWGYDP	Bertoldo *et.al.*, 1999
Pullulanase I	*Thermus sp.*	DAVYNH	DGFRFDLMGV	EGWD	YVECHD	YNWGYDP	Tomiyasu *et.al.*, 2001
Pullulanase I	*K.aerogenes*	DVVYNH	DGFRFDLMGY	EGWD	YVSKHD	YNWGYDP	Katsuragi *et.al.*, 1987
Pullulanase I	*T.maritima*	DMVFPH	DGFRFDQMGL	EPWG	YAACHD	YNWGYDP	Bibel *et.al.*, 1998
α-amylase	*T.maritima*	DLVINH	DGFRIDAAKH	EVFS	LENHD	-	Liebl *et.al.*,1997
α-amylase	*B.megaterium*	DLVVNH	DGFRLDAALH	EVWD	LTNHD	-	Lu *et.al.*, 1991
α-amylase	*D.thermophilum*	DLVVNH	DGFRLDAAKH	EVWD	LRHND	-	Fukysumi *et.al.*, 1988
Neopullulanase	*B.stearothermophilus*	DAVFNH	DGWRLDVPNE	EIWHD	LLGSHD	-	Hondoh *et.al.*, 2003
Neopullulanase	*T.vulgaris*	DAVFNH	DGWRLDVANE	QIWHD	LLDSHD	-	Tonozuka *et.al.*, 1993
		* * *	** * *	*	**	*******	

1.3.4 Pullulanase

Pullulanase (EC 3.2.1.41) is also known as pullulan-6-glucanohydrolase. Pullulanases can be categorized into 4 different groups based on different substrate specificities (Table 1.3).

Pullulanase of *Thermopallium natronophilum* has been identified as pullulanase type I. Most of the pullulanases characterized to date have been type II pullulanases. According to the literature available, pullulanase type I has been characterized in *Klebsiella aerogenes* (Takizawa and Murooka, 1985), *Klebsiella pneumoniae* (Pugsley *et al.*, 1986), *Pyrococcus weosei* (Rüdiger *et al.*, 1995), *Bacteroides thetaiotaomicron* 95-1 (Abdullah and French, 1966), *Bacillus* sp. S-1 (Lee *et al.*, 1997), *Bacillus acidopullulyticus* (Frantzen and Svendsen, 2002), *Bacillus flavocaldarius* KP 1228 (Suzuki *et al.*, 1991), *Bacillus thermoleovorans* US105 (Messaoud *et al.*, 2002), *Bacillus* sp. KSM-1876 (Hatada *et al.*, 2001), *Thermus* Strain IM6501 (Kim *et al.*, 2000), *Thermus thermophilus* HB8 (Tomiyasu *et al.*, 2001), *Thermus aquaticus* YT-1, *Thermus caldophilus* GK-24 (Kim *et al.*, 1996), *Caldicellulosiruptor saccharolyticus* (Albertson *et al.*, 1997), *Thermotoga maritima* (Kriegshauser and Liebl, 2000), and *Fervidobacterium pennavorans Ven5* (Bertoldo *et al.*, 1999). To date, there are still no published structural studies of type I pullulanase, although crystallization of type I pullulanase of *Fervidobacterium* has been carried out by Lebbink *et al.* (2000). Instead, the structure of neopullulanase (EC 3.2.1.135) annotated in Fig. 1.8 from *Bacillus stearothermophilus* has been solved by Hondoh *et al.* (2002).

In *Fervidobacterium pennavorans Ven*5, type I pullulanase is encoded by the *pulA* gene. This gene is found to have 50.1% homology with *pulA* gene of *Thermotoga maritima*. (Bertoldo *et al.*, 1999).

Pullulanase type I specifically hydrolyses α-1,6-linkages in amylopectin, pullulan, or limit dextrins. This enzyme requires at least two α-1,4 linkages in the surrounding area of α-1,6 linkages for hydrolysis to occur. This enzyme has been found not to be

able to hydrolyse substrates such as dextran and isomaltotriose, which are composed strictly of α-1,6 linkages (Kainuma *et al.*, 1978).

Pullulanase Type I contains the four-conserved regions that define members from the α-amylase family. There is an additional conserved region that is only found in type I pullulanase and it is a region of seven amino acids i.e. YNWGYNP. This region is found in type I pullulanase from *F.pennavorans Ven5, T.maritima, Thermus sp.* and *C.saccharolyticus*. This highly conserved region was speculated to be a region that is involved in substrate binding or catalytic activity (Bertoldo and Antranikian, 2002).

Table 1.3: The four types of pullulanase and their substrate specificities.

Type of Pullulanase	Substrate Specificities
Type I pullulanase	Hydrolyses α-1,6 glycosidic bonds in pullulan or branched oligosaccharides, forming maltotriose units (Fig.1.7) and linear oligomers, respectively.
Type II pullulanase /amylopullulanase	Hydrolyses both α-1,4 and α-1,6 glycosidic linkages.
Pullulan hydrolase type I/ neopullulanase	Hydrolyses pullulan to produce a mixture of end products, namely panose, maltose and glucose (Fig.1.7) in a final molar ratio of 3:1:1. It hydrolyses α-1,4 and α-1,6 glycosidic linkages and also catalyses transglycosylation reactions to form α-1,4 and α-1,6 glycosidic linkages.
Pullulan hydrolase type II/isopullulanase	Hydrolyses α-1,4 bonds in pullulan generating end product isopanose units (Fig.1.7).

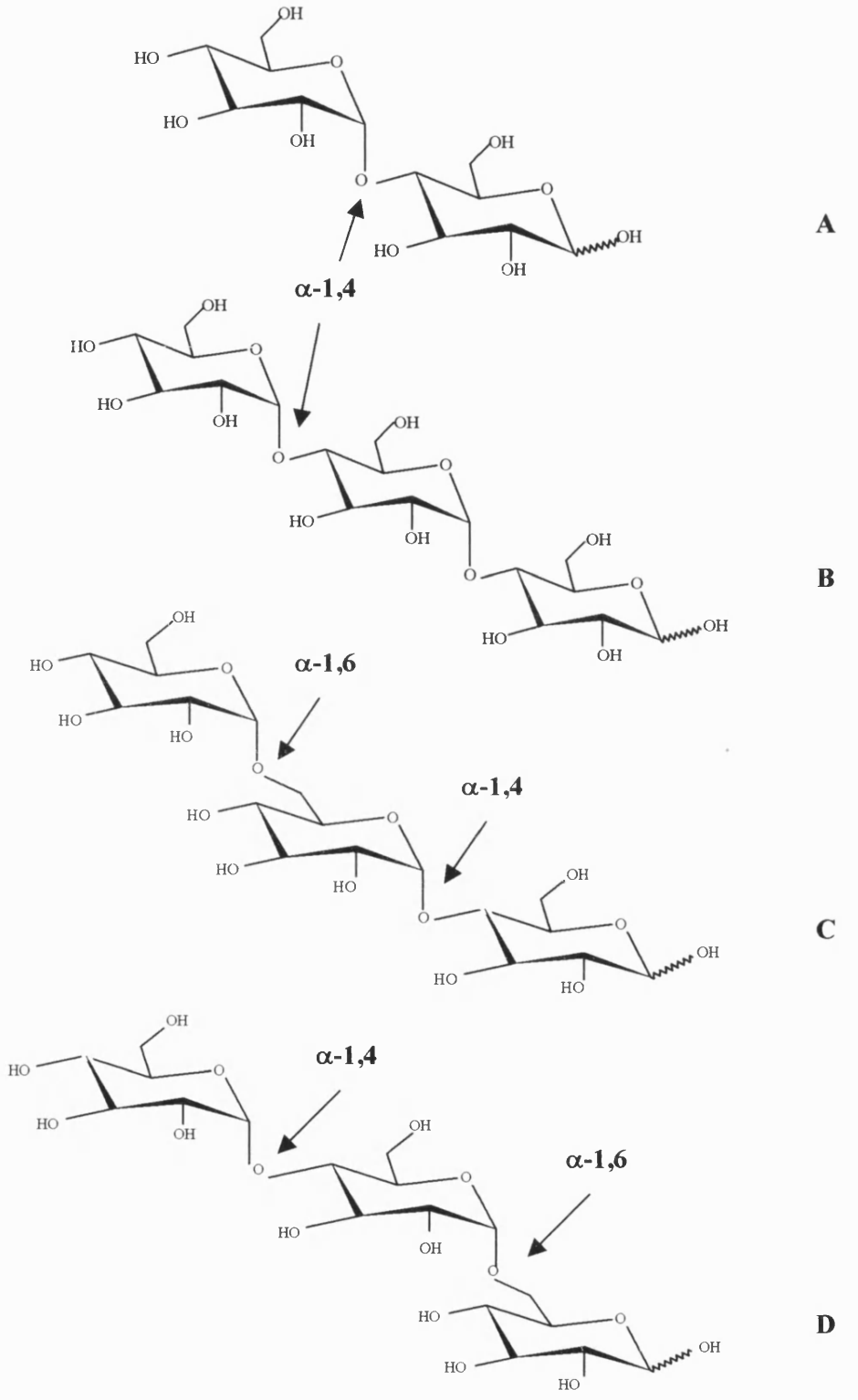


Fig.1.7: Chemical structures of maltose (A), maltotriose (B), panose (C) and isopanose (D). Adapted from Aoki *et al.* (1997).

Fig.1.8: (a) The monomeric structure of the *B.stearothermophilus* neopullulanase (Hondoh *et al.*, 2003). A ball and stick model has been used to identify the three catalytic residues, Asp328, Glu357 and Asp424. The orange sphere is the calcium ion. The active sites in the monomer are shallow and wide. Domains A, B and N of both monomers will come together to give active sites that are narrower in the dimer.
(b) The dimer annotated as Mol –1 and Mol-2 coming together. Bound substrates on both catalytic sites are panose molecules.

1.4 Substrate

1.4.1 Starch

Starch is composed of two main components i.e. 15%-25% amylose and 75%-85% amylopectin. Amylose is composed of glucose units linked together by α-1,4 glycosidic linkages while amylopectin is a branched polymer with α-1,4 bonds linking the glucose monomers and α-1,6 bonds are found at the branched points. Due to the complexity of the structure of this macromolecule, a combination of enzymes is normally used in the hydrolysis of this polymer in industry (Niehaus *et al*, 1999). The chemical structure of starch is illustrated in Fig.1.9.

1.4.2 Pullulan

Pullulan was first isolated and characterised from *Aureobasidium pullulans* by Bernier in 1958 and was named pullulan by Bender *et al.* in the following year. Pullulan is composed of glucopyranose units linked together by α-1,4 glycosidic linkages into maltotriose units which are in turn join together by α-1,6 glycosidic bonds (Tarabasz-Szymanska, 1999). Pullulan can be hydrolysed completely by the enzyme known as pullulanase type I. Pullulan with its specific structure has made it a source of maltotriose and as a substrate for the study of pullulanase activity (Tarabasz-Szymanska, 1999). The chemical structure of pullulan is illustrated in Fig. 1.10.

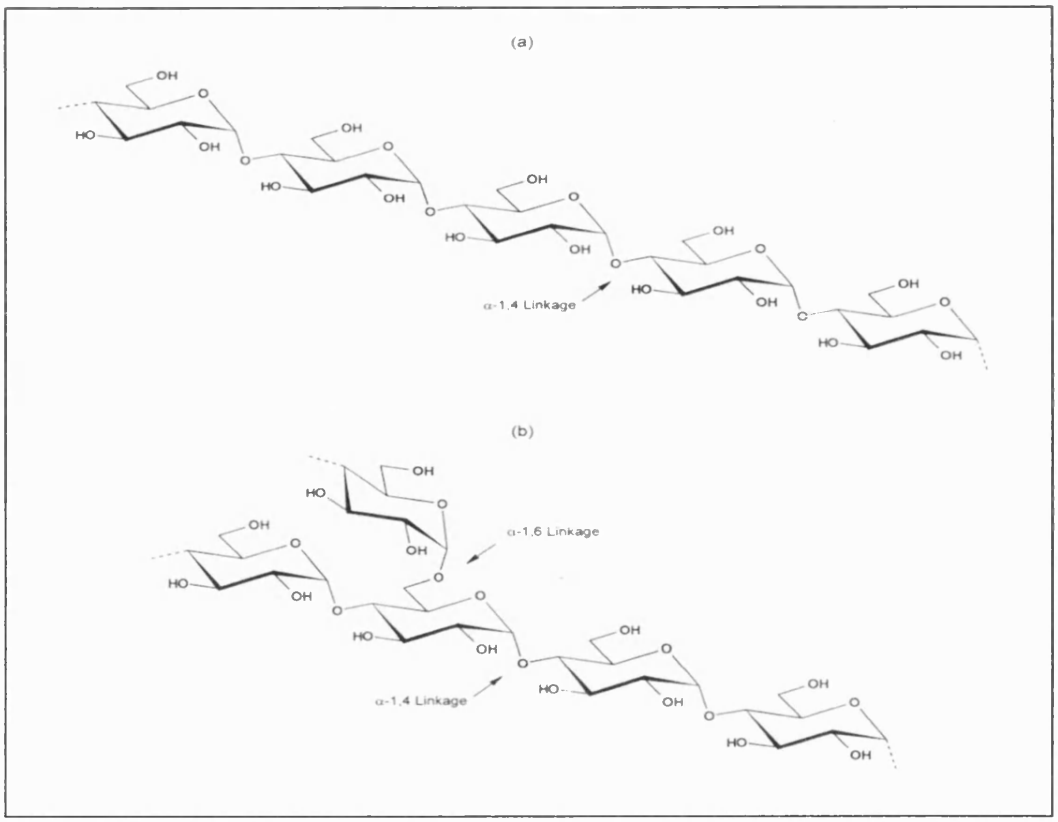


Fig 1.9: (a) Amylose: glucose monomers linked by α-1,4 bonds
(b) Amylopectin: glucose monomers linked by α-1,4 bonds with α-1,6 bonds at branch points (Thompson, 1998).

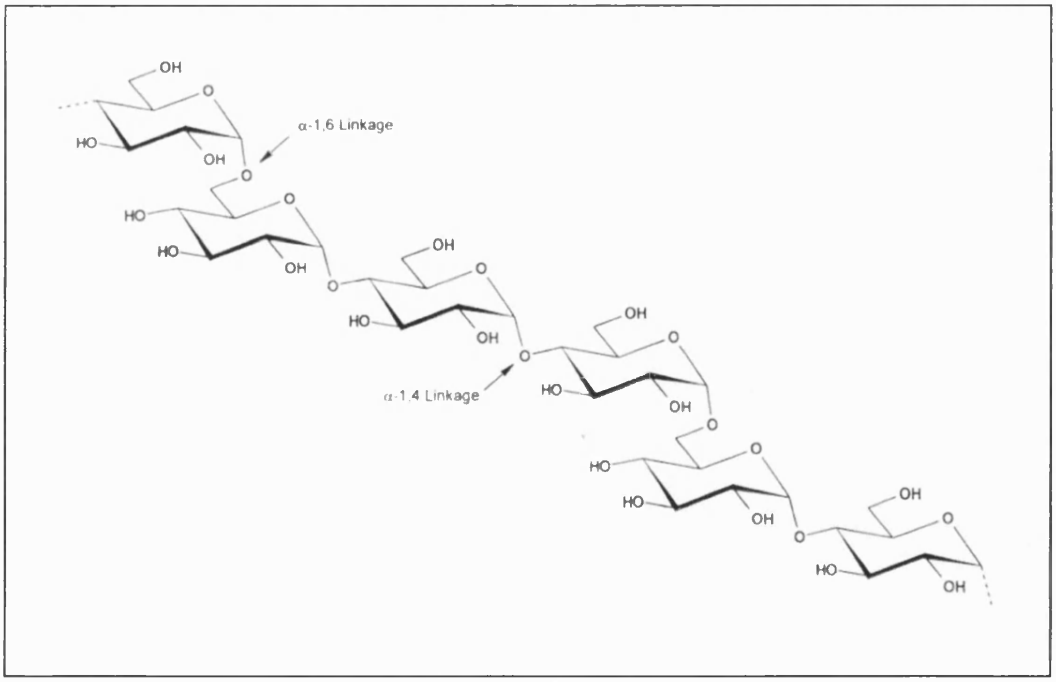


Fig 1.10: Structure of pullulan: maltotriose units linked together by α-1,6 bonds (Thompson, 1998).

1.5 Mode of Hydrolysis

The hydrolysis of a glycosidic bond either leads to an inversion or retention of the anomeric configuration of the sugar, and in both mechanisms a pair of carboxylic acids and oxo-carbonium-ion like transition states are involved (Rye, C.S. and Withers, S.G., 2000). Enzymes of the α-amylase family catalyse the hydrolysis of starch without changing the anomeric configuration of the sugar. Hydrolysis proceeds via a double displacement mechanism, whereby the first step involves a general acid catalyst causing the protonation of the glycosidic oxygen. This in turns creates a transitional step in the form of an oxo-carbonium ion, which then collapses into an intermediate covalently linked to a nucleophile. In the second step, the covalently linked glycosyl enzyme intermediate is then attacked by a water nucleophile aided by the base form of the acid catalyst and hydrolysis occurs (Veen *et al.*, 2000; Uitdehaag *et al.*, 1999; Davies and Wilson, 1999). A diagrammatic representation of the process is illustrated in Fig.1.11.

As mentioned in the previous section 1.3.3, three carboxylic acid side chains (two aspartic acids and one glutamic acid) are involved in catalytic activity in α-amylase. These three residues have been identified as Asp206, Glu230 and Asp297 in α-amylase from *A.oryzae* (Matsuura *et al.*, 1984) and Asp229, Glu257 and Asp328 in CGTase from *B.circulans* (Veen *et al.*, 2000).

In the case of CGTase from *B.circulans*, Asp229 is the nucleophile in the first step of the reaction, generating the intermediate. This nucleophile then serves as a leaving group in the second step of the hydrolysis. Glu257 on the other hand serves as the acid/base catalyst in which it serves as an acid during the glycosylation reaction but will have to switch role to be a general base as deglycosylation takes place. Through experimental studies, only a pair of carboxylic acids is involved in the double displacement reaction. In *B.circulans*, Asp328 has been proposed to be involved in substrate binding (Veen *et al.*, 2000).

A more detailed discussion of these conserved residues can be found in Chapter 5.

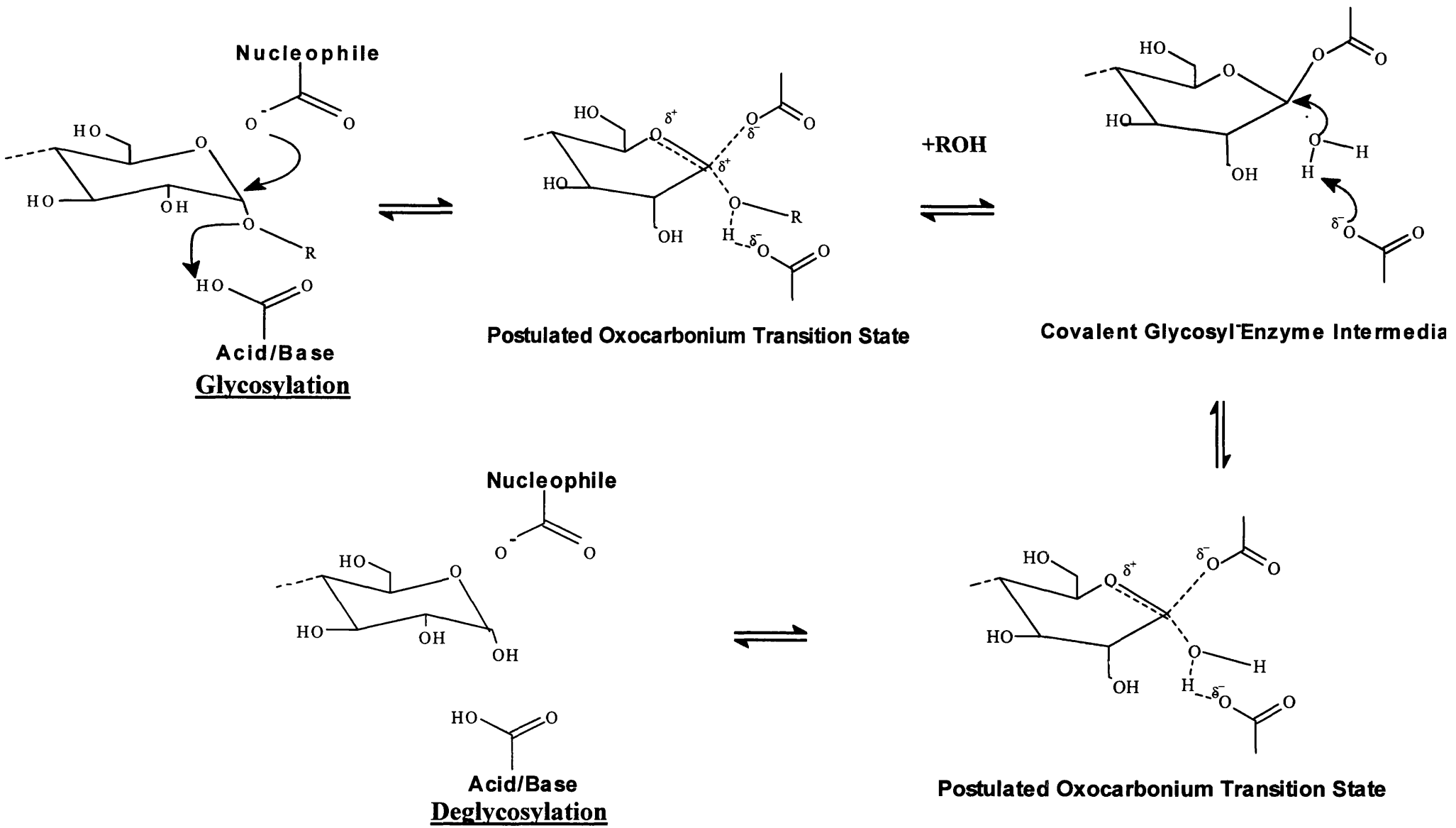


Fig.1.11: Reaction mechanism of the α-amylase family. This is adapted from Veen *et.al.*, (2000) and McIntosh *et.al.* (1996). This figure describes the hydrolysis of an α-1,4 glycosidic bond.

1.6 Colourimetric Assay for Pullulanase Type I

The enzyme assay used to measure type I pullulanase activity is based on detecting the presence of reducing sugars i.e. the free carbonyl group (C=O) (Bernfeld, 1995). This method is based on the oxidation of the free carbonyl group and the reduction of 3,5-dinitrosalicylic acid (DNSA) to 3-amino, 5-nitrosalicylic acid (ANSA) (Fig.1.12). This is a discontinuous assay and the product can be measured at 550nm using the spectrophotometer. This reaction requires four protons, 100°C heat and high pH for the colour change to occur. One unit of enzyme activity was defined as the amount of enzyme producing 1 μmole of reducing sugars per min under assay conditions.

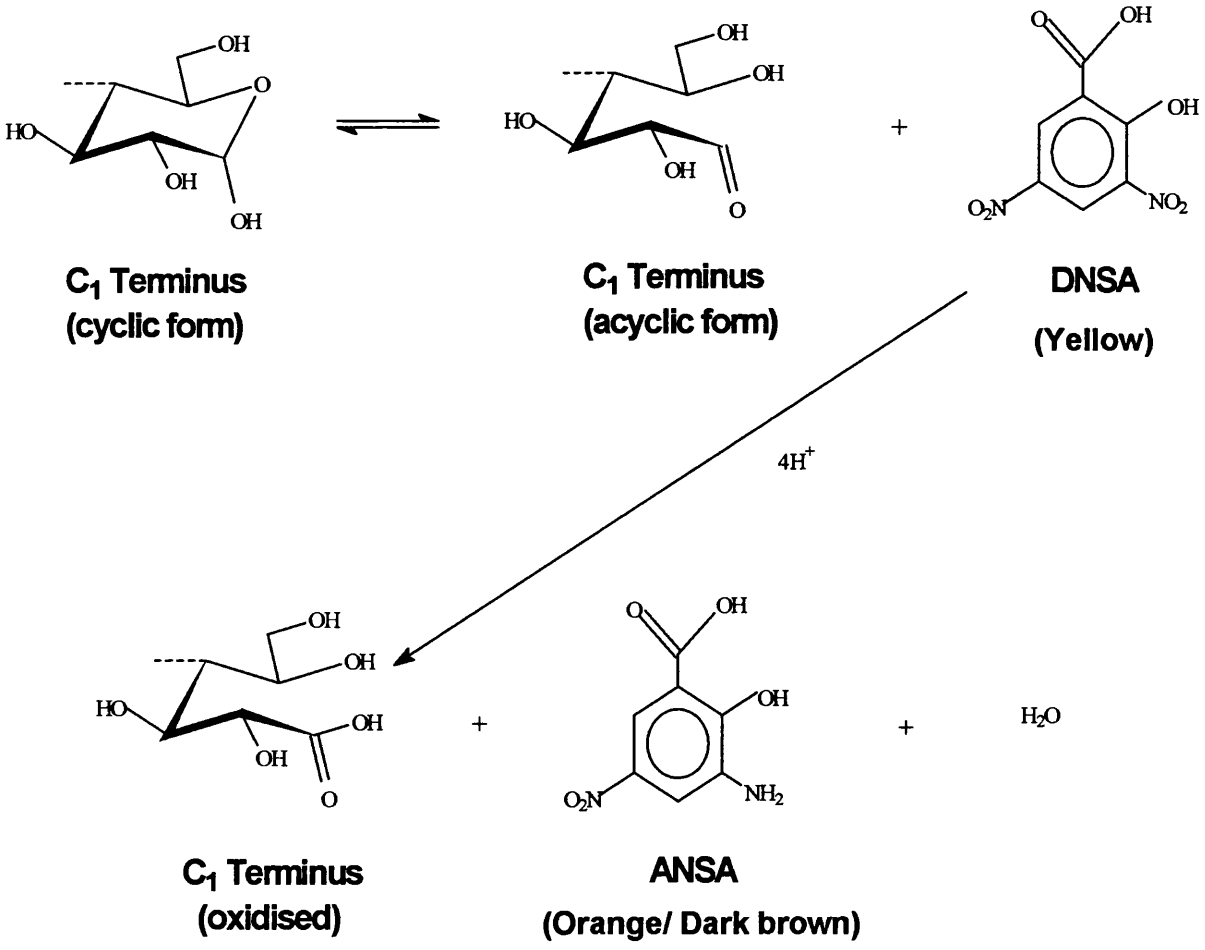


Fig.1.12: The principles of the enzyme assay used to detect the reducing sugar liberated by pullulanase type I of *T.natronophilum* (adapted from Thompson, 1998).

1.7 Industrial Enzyme Applications

Industrial enzymes had a global market of USD1.5 billion in 2000. There was an increase of USD0.5 billion from year 1995 to 2000. Generally, industrial enzymes are categorised into three sections i.e. food enzymes, animal feed enzymes and technical enzymes. Technical enzymes are enzymes that are widely used in the detergent, textile, starch and baking industries. They are the most profitable commodity in industrial enzymes (Krik *et al.*, 2002).

Extremophilic microorganisms are often of interest to both academia and industry due to their unique survival abilities and the use of their enzymes in various industries. The various industries that exploit extremophilic enzymes are listed in Table 1.4.

In this section, the potential application of the type I pullulanase from *T.natronophilum* will be discussed.

Table 1.4: Extremophilic enzymes and their uses in various industries (adapted from Schiraldi and De Rosa (2002) and Novozyme website (www.novozymes.com).

Extremophilic Enzyme	Applications
Amylases, Glycosidases	Starch processing (OPTIMAX™ HI-DEX, Genencor), baking industry (Multifresh®, Genencor), textile industry (Deniprime®, Genencor), detergent industry (Termamyl® Ultra, Novozyme)
Lipase	Paper industry (Resinase® HT, Novozyme), detergent industry (Lipex®, Novozyme)
Xylanases	Pulp and paper industry (bleaching) (MULTIFECT® XYLANASE, Genencor)
Proteases	Food processing (FUNGAL PROTEASE 500,000, Genencor), detergent industry (Savinase® Ultra, Genencor)
DNA polymerases	Genetic Engineering (Taq DNA Polymerase)
Cellulases	Detergent industry (Denimax® 601L, Novozyme)

1.7.1 Detergent Industry

Detergents are used in household laundry, dish washing machines and also in industrial and institutional laundry. Detergents can be biological or non biological. In biological detergents, enzymes are used. The four major enzymes used at present are lipases, cellulases, proteases and amylases (www.novozymes.com/cgi-bin/bvisapi.dll/portal.jsp).

Most detergents are able to remove 95% of a particular stain but the remaining 5% can behave as invisible glue, which will attract dirt, and this is particularly true for starch stains. Thus, in order to have 'total cleaning', a terminology frequently coined in the detergent industry, enzymes have to be used (www.novozymes.com/cgi-bin/bvisapi.dll/portal.jsp).

Carbohydrates are part of our daily diet. For complete removal of carbohydrate stains on plates and utensils, automatic dishwashing detergents (ADDs) with α-amylase and pullulanase have to be used. Starch stains on plates behave like starch stains on clothes where eventually they behave as glue and attract dirt, leaving a dull appearance on plates (www.novozymes.com/cgi-bin/bvisapi.dll/portal.jsp).

Detergent enzymes are the most profitable of technical enzymes. Due to the nature of all detergents, particularly the high pH and the high temperature conditions in which they operate, it is important to use enzymes that can withstand these conditions.

Type I pullulanase of *T.natronophilum* is a starch debranching enzyme. Considering the fact that this enzyme is isolated from an organism that is both thermophilic and alkaliphilic, there is a particular interest in exploiting this enzyme in the detergent industry.

1.7.2 Textile Industry

Cotton or cotton synthetic fibres are often coated with strengthening agents such as starch in order to prevent breaking during the weaving process. This process is known as sizing. 75% of size used for textile industries is starch.

Following the production of fabric, further downstream processing such as bleaching or dying is carried out. These downstream processes are known as the wet processing in the industry, and starch granules left on fabric from the weaving process have to be removed. Starch removal treatment is known as desizing. There are several ways in which desizing can be carried out and they are listed in Table 1.5.

Table 1.5: Types of sizing agents and their advantages and disadvantages.

Types of Desizing	Pros and Cons
Detergents	Only when size used is water soluble
Oxidative chemicals (e.g. persulphate/ bromide)	Damages the cotton fibre
Enzyme (e.g. amylases, pullulanases)	Does not damage cotton fibres. Only used when size is starch.

There are three stages to the desizing process: impregnation, incubation and after wash. Prior to the impregnation stage, the fabric is washed with water to remove any water-soluble contaminants. After, the fabric is heated to gelatinise the starch in order to facilitate enzyme-substrate interactions. This impregnation stage is normally carried out at temperatures above 70°C. Following this, the fabric is incubated with enzyme for 2-16 hours depending on the enzyme stability and enzyme activity. The iodine test is then used to detect the presence of residual starch. Once starch is not present, the fabric is then washed with detergents, sodium hydroxide and water to remove the enzymes and hydrolysis products.

Due to the long incubation time at elevated temperature, it is crucial that thermo-stable hydrolases with acceptable enzyme activity are employed (http://science.ntu.ac.uk/research/EnzyTex/EnzRep2.html#2.2.1.).

1.7.3 Baking Industry

In the baking industry, hydrolases are often added to bread dough to prevent the recrystalisation of starch. Pullulanase is often added with α-amylase to retard staling. Amylopectin in starch can be broken down by pullulanase and this in turn will help the bread to retain its softness, thus extending its shelf life. (www.foodproductdesign.com). During the baking process, high heat is often employed. Thermo-stable hydrolases will not be denatured so rapidly and thus are able to improve the quality of the bread.

Chapter 2: Materials and Methods

2.1 Molecular Biology Materials and Methods

2.1.1 Enzymes, reagents and other materials

All chemicals used are from Sigma Aldrich (Dorset, UK), and all solvents are from Fisher Scientific U.K. Limited (Loughborough, Leicestershire, UK) unless otherwise specified. All restriction enzymes are from NEB (Hertfordshire, UK) unless otherwise specified. All primers used in PCR are from MWG (Germany).

2.1.2 Media

2.1.2.1 LB Medium

Table 2.1: Ingredients for LB media.

Ingredients	Weight (g)
NaCl	10g
Tryptone	10g
Yeast Extract	5g
Final Volume	1L (pH7.0)

To make 1L of LB agar, 20g of agar was added to 1L of LB medium before autoclaving.

2.1.2.2 SOC Medium

Table 2.2: Components used to make SOC medium.

Ingredients	Weight or Volume
MQH_2O	950ml
Tryptone	20g
Yeast Extract	5g
NaCl	0.5g
250mM KCl	10ml
2M $MgCl_2$	5ml
Adjust pH to 7.0 with 5M NaOH	
Adjust volume to 1L, Autoclave	
Add 20ml of filter sterilised 1M Glucose	

2.1.3 gDNA extraction from *T. natronophilum*

0.5g of cell paste was first freeze-thawed thee times to disrupt the cell toga and cell membrane. The cell paste was then resuspended in 9.5ml of 10× TE buffer (100mM Tris; 10mM EDTA, pH 8). 0.5 ml of 10% SDS and 50µl of 20mg/ml proteinase K were then added. The solution was then mixed and incubated at 50°C for 1 h. Following this, 8ml of 5M NaCl was added and the solution was mixed gently. Then, 1.5ml of 10% CTAB/0.7M NaCl solution was added. It is crucial to keep the concentration of NaCl above 0.7M in order to keep DNA soluble. Upon addition of CTAB/NaCl, the solution was mixed gently and incubated at 65°C for 20 min.

An equal volume of chloroform/isoamyl alcohol (24:1) was then added and the sample was spun at 12k rpm for 20min. A white precipitate separating two aqueous

phases would be present and the upper aqueous phase, which contains the DNA, was then carefully removed for further extraction. This extraction step was repeated two or three times till the white precipitate separating the phases was no longer present. Following this, an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added. Samples were spun at 12000× rpm for 20 min and the upper aqueous phase was carefully removed and kept. In order to remove the phenol in the DNA sample, an equal volume of chloroform was added. The sample was spun at 12000× rpm for 20 min and the upper aqueous phase was removed.

To precipitate the DNA, two volumes of ice-cold ethanol with respect to sample volume were added. DNA was then spooled out with a heat-sealed Pasteur pipette and redissolved in 1ml of MQH_2O. To facilitate the solubilisation of DNA, the sample was kept on ice for 30 min. To degrade RNA present in the sample, 10µl of 0.05mg/ml RNase A was added and sample was incubated at 37°C for 1 h. A final round of chloroform/isoamyl alcohol (24:1) precipitation was performed to remove RNase. DNA was once again precipitated with two volume of ice-cold absolute ethanol, spooled and resuspended in 70% ethanol. The sample was then spun at 13000× rpm for 5 min to pellet the DNA. 70% ethanol was then aspirated off carefully. The sample was washed twice in 70% ethanol and air-dried for 10-15 min before redissolving the pellet in 500µl of Tris/HCl pH 8.0. The sample was left at 4°C overnight to facilitate DNA solubilisation (Moore, 1994).

2.1.4 Quantification of DNA

10 µl of gDNA was added into 500 µl of 1× TE buffer (10mM Tris pH8, 1mM EDTA). The sample was vortexed for 1 min and absorbance readings were taken at A_{260} and A_{280}. Nucleic acid without phenol and protein contamination should give A_{260}: A_{280} ratios of ~1.2. At A_{260}, an OD reading of 1 contains ~50µg/ml of double stranded DNA (Sambrook *et al.*, 2001). A 1µl sample was also run on a 0.8% agarose gel alongside a high DNA mass ladder to quantify the DNA present.

2.1.5 Agarose gel electrophoresis

DNA samples were analysed using agarose gel electrophoresis. 0.7% (w/v) Agarose gel (Helena Biosciences, UK) separates DNA in the range of 800-12,000bp. For the preparation of an agarose gel, the appropriate amount of agarose was dissolved in 100ml of 1× TAE buffer (40mM Tris-acetate, 1mM EDTA, pH8) by heating. When the gel mixture cooled to 45°C, ethidium bromide was added to a final concentration of 0.5mg/ml and immediately poured into a perspex gel cast. A comb was placed into position to create the wells and the gel was then left to set at room temperature. Once set, the gel was put into an electrophoresis tank and immersed in 1× TAE buffer. Prior to loading, the comb was removed. Samples were mixed with the 6× loading dye (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol *ff*, 40% (w/v) sucrose) in a 1:1 ratio. Once loaded, the gel was electrophoresed at a constant voltage of 100V and DNA bands were visualised using a UV transilluminator.

Table 2.3: The percentage of agarose used in experiments and the degree of separation of these gels.

Agarose Gel (%)	Range of Separation (bp)
0.5	1,000-30,000
0.7	800-12,000
1.0	500-10,000

2.1.6 Restriction digestion of DNA

DNA digestion with restriction endonucleases was carried out as specified in the manufacturer's manual. The DNA was incubated for 1 h or overnight at 37°C with the enzyme in the appropriate buffer, which is supplied as a 10× stock by the manufacturer. For double digestion of DNA, a suitable common buffer, which gives maximal activity to both enzymes, was used. If double digestion is not feasible, the DNA was first digested with one enzyme in its appropriate buffer and the DNA was purified via agarose gel extraction before the second enzyme was used.

2.1.7 Polymerase Chain Reaction: Amplification of DNA

The polymerase chain reaction has been used to amplify DNA. To verify that the required fragments were amplified, single primer controls were set up.

A typical PCR reaction is set up as described in Table 2.4. Unless otherwise stated, all PCR reactions were carried out using $Vent_R$® DNA Polymerase (NEB, UK). 10× reaction buffer supplied by the manufacturer was used as a 1× concentrate. 1× reaction buffer contains 10mM KCl, 10mM $(NH_4)_2SO_4$, 20mM Tris-HCl (pH 8.8, at 25°C), 2mM $MgSO_4$, 0.1% Triton X-100. dNTP set (Bioline GmbH, Luckenwalde, Germany) supplied as dATP, dTTP, dCTP and dGTP at 100mM each was used to make dNTP stock with a final concentration of 20mM of each nucleotide. PCR reactions were carried out in 0.5ml tubes. The thermal cycler used was a Ceteus DNA Thermal Cycler (Perkin-Elmer, Norwalk CT, USA) or an Eppendorf Mastercycler (Eppendorf, Germany). For the former thermal cycler, an additional 20µl of mineral oil was added to the PCR reaction to prevent evaporation.

PCR was carried out with a hot start of 94°C for 3 min then followed by 30 cycles of amplification with denaturation at 96°C, annealing at temperature 1°C lower than the Tm of primers and elongation at 72°C at 1kb per min extension time. For analysis, 10µl of PCR samples were then loaded onto a 0.7% agarose gel.

2.1.7.1 Nested PCR

Nested PCR is similar to the conventional PCR described in Table 2.4. Nested PCR involves a sequence of two PCR reactions with different sets of primers. The forward primer for the second set PCR will be downstream of the forward primer in the first PCR, whilst the reverse primer of second round PCR will be upstream of the reverse primer of first round PCR. Thus, only the required PCR products amplified by the first set of primers can be amplified in the second round of PCR.

Table 2.4: Components to set up the standard PCR.

	Double Primer Reaction	Single Primer Control (Forward Primer)	Single Primer Control (Reverse Primer)
Template DNA (50-100ng/μl)	1μl	1μl	1μl
10X Reaction Buffer	5μl	5μl	5μl
Vent$_R$® DNA Polymerase (2U/μl)	1μl	1μl	1μl
dNTP(20mM each)	1μl	1μl	1μl
Forward Primer (100pmol/μl)	1μl	1μl	-
Reverse Primer (100pmol/μl)	1μl	-	1μl
MQH$_2$O	40μl	41μl	41μl

2.1.7.2 Inverse PCR

Inverse PCR was used to investigate the unknown sequences adjacent to known DNA sequences. This method involves the digestion of 1-2 μg of gDNA with a restriction enzyme and later religation of the fragments to form intramolecular, recircularised DNA. The specific primers used were designed to move in opposite direction reading into the unknown gene sequence. Several restriction enzymes were used simultaneously. After digestion, the DNA was ethanol precipitated and resuspended with MQH$_2$O in 50× the volume of the undigested DNA used initially. Only a fifth of this was then used in ligation. This dilution step was to encourage intramolecular

religation. T4 DNA ligase (400U/μl) from Promega, UK was used for ligation. Ligation was set up according to the manufacturer's instruction and was left overnight for optimum results. This was then followed by ethanol precipitation of the sample. The sample was resuspended in the same volume as the digestion reaction and a fifth of this is used in PCR reaction.

2.1.7.3 Colony PCR (screening of library)

Colony PCR was used in screening of the gDNA library for potential pullulanase gene clones. This involves taking a bacterial colony and resuspending the colony in 100μl of MQH_2O. The suspension was then boiled at 100°C and the lysed cell suspension was then spun down at 13,000 ×g for 1min. 5μl of the supernatant was then used as template for PCR.

2.1.8 Purification of PCR products

The desired PCR bands were excised from agarose gel purified with QIAEXII DNA Extraction Kit (Qiagen, Germany) or QIAquick PCR Purification Kit (Qiagen, Germany) as per manufacturer's instructions.

2.1.9 Miniprep

10ml of cell culture was pelleted and plasmid was extracted using Nucleospin® Plus Minipreps Kit (BD Biosciences, Oxford, UK) as per manufacturer's instructions.

2.1.10 Dephosphorylation of plasmid DNA

Dephosphorylation of 5'phosphate ends of DNA was carried out to prevent self-ligation of linearised plasmid DNA. This enzymatic reaction was carried out using shrimp alkaline phosphatase (Roche, UK). Approximately 1μg of linearised DNA was incubated with 2U of shrimp alkaline phosphatase in a 1× buffer provided by the manufacturer (10× concentrate) at 37°C for 1 h. The whole DNA sample was then run on agarose gel and the appropriate sized band was excised and extracted. Shrimp alkaline phosphatase can be used together with *BamHI* during digestion of DNA with this enzyme. To check the efficiency of dephosphorylation, the plasmids were ligated and transformed into JM109 heat shock competent cells.

2.1.11 DNA Ligation

Ligation of cohesive ends was carried out using T4 DNA Ligase. The ligation reaction is normally done in a 10μl or 20μl volume. For a 10μl reaction, 1U of T4 DNA ligase was used. Vector: insert ratios of 1:1, 1:2 and 1:3 were used. The estimation of vector and insert used can be calculated with the following formula:

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{molar ratio of insert: vector (e.g. 3/1)} = \text{ng of insert}$$

2.1.12 Ethanol precipitation of DNA

The volume of sample was first estimated. Then, an appropriate amount of 3M sodium acetate, pH 5.2 was used to give a final concentration of 0.3M sodium acetate. Following this, two volumes of ice-cold ethanol were added. The solution was mixed and left on ice for 30min to facilitate DNA precipitation. The sample was then spun at 14,000 ×g at 4°C for 30 min. The supernatant was then removed carefully and the pellet was washed twice with 70% ethanol. The sample was then air dried and resuspended in an appropriate volume of 5mM Tris-HCl, pH 8.5. For sequence analysis, the DNA sample was resuspended in autoclaved MQH_2O.

2.1.13 DNA sequencing

DNA sequencing was carried out by Paul Jones from the Biology and Biochemistry Department, University of Bath, using equipment based on the method described by Sanger *et al.* (1974). Each ddNTP was labelled with four different fluorescent dyes and DNA was analysed on a polyacrylamide gel using an ABI PRISM^{TM} 377 DNA sequencer (Applied Biosystems, Fostercity CA, USA). Samples were prepared in a final volume of 6µl with 30-90ng of DNA and 10pmol of sequencing primer. Sequences obtained were analysed using Chromas Version 1.45 (Technelysium Ltd, Australia) and aligned using PILEUP in GCG® Wisconsin Package® (Accelrys Inc., Cambridge, UK). The MSF file produced from the multiple sequence alignments were then viewed and presented in GeneDoc Ver2.6.001 (Nicholas *et al.*, 1997).

2.1.14 Preparation of Heat Shock Competent Cells

A single colony of *E.coli* JM109 was inoculated into a 2.5ml LB culture. The culture was incubated overnight at 37°C at 225 rpm. The entire inoculum was then used to inoculate 250ml LB media containing 20mM $MgSO_4$, grown in a 1L flask for better aeration. When the OD reached 0.6, cells were harvested by centrifugation at 4,500 ×g at 4°C for 5 min. The cell pellet was then resuspended in 100ml of ice-cold TFB1 (Table 2.5). Cells were then kept on ice for 5 min and pelleted at 4,500 ×g for 5 min at 4°C. The supernatant was then removed and pellet was resuspended in 10ml of ice-cold TFB2 (Table 2.5). The cells were then kept on ice for 1 h and aliquoted in 200µl volumes in 0.5ml tubes. Cells were snap-frozen in a dry ice/ethanol bath and stored at –80°C.

Table 2.5: Components used to make TFB1 and TFB2 stock solutions.

Solutions	Ingredients
TFB1	30mM $C_2H_3KO_2$, 10mM $CaCl_2$, 50mM MnCl, 100mM RbCl, 15% glycerol (v/v), pH adjusted to 5.8 with 1M CH_3COOH, filter-sterilised).
TFB2	10mM MOPS, 75mM $CaCl_2$, 10mM RbCl, 15% (v/v) glycerol, pH adjusted to 6.5 with 1M KOH, filter-sterilised.

2.1.15 Transformation of DNA into bacteria

2.1.15.1 Heat Shock Transformation

E.coli JM109 competent cells were first thawed on ice. Approximately 50ng of DNA was mixed with 50µl of competent cells. The cells were then incubated on ice for 30 min, followed by a heat shock at 42°C for 1.5 min. Following this, cells were incubated on ice for 5 min and 0.95ml of LB containing 20mM glucose, were added. The cells were incubated at 37°C for 1 h with agitation at 170 rpm. 100µl of the culture was then plated on LB plates with the appropriate antibiotic.

2.1.15.2 Electroporation

Electroporation was carried out using electroporation cuvettes (Bio-Rad Laborotaries GmbH, München, Germany) and a MicroPulser^{TM} (Bio-Rad Laborotaries GmbH, München, Germany) as per manufacturer's instructions. All 1.5ml tubes and cuvettes were pre-chilled on ice before electroporation was carried out.

2.1.16 Selection of Transformed Cells

2.1.16.1 Antibiotic Selection

The plasmids and their appropriate antibiotics described throughout this thesis are annotated in Table 2.6. 100µl of transformed cells was plated on LB plates with 100µg/ml of the appropriate antibiotics and incubated at 37°C in an incubator.

Table 2.6: Plasmids used in this project and their appropriate antibiotics.

Type of Plasmid	Antibiotic (100µg/ml)
pGEMT, pUC18, pTrueBlu*Rop*	Carbenicillin
pET28b	Kanamycin

2.1.16.2 Blue/White Selection

100µl of 10mM IPTG (23.8mg in MQH_2O, filter sterilised) was first added to 100µl of SOC medium. The solution was mixed well before another 100µl of 1% (w/v) X-gal was added. 1% X-gal was prepared by dissolving 0.01g of X-gal in 10ml of N,N'-dimethylformamide. All 300µl was then plated on an LB plate with 100µg/ml of the appropriate antibiotic and air dried in the sterile hood. When the plate was dry, 100 µl of transformed competent cells was then plated.

2.2 Protein Materials and Methods

2.2.1 Enzymes, Reagents and Other Materials

Acetic acid, butanol, calcium chloride, EDTA, absolute ethanol, glucose, hydrochloric acid, methanol, sodium chloride, sodium hydroxide and Trisma Base were of the highest quality and were obtained from Fisher Scientific U.K. Limited (Loughborough, Leicestershire, UK).

Acrylamide/Bis-acrylamide mixture, ammonium persulphate, pullulan, starch, amylose and amylopectin, were supplied by Sigma-Aldrich (Dorset, UK). Protein molecular weight standards and Bradford protein estimation reagents were supplied by Biorad, Hercules CA, USA.

Enzyme assays were measured using a CARY 300 BIO spectrophotometer (Varian Ltd., Walton-on-Thames, UK).

2.2.2 Protein Estimation with Bradford Assay

Protein estimation was carried out using a method based on Bradford (1976). The final volume of each assay is 1ml. BSA was used as a standard and a series of different BSA concentrations from 1-100μg with a 10μg increment was prepared. The different BSA concentrations were made up to a final volume of 800μl with MQH_20. 200μl of Bradford reagent (0.01% (w/v) Coomasie brilliant blue G-250, 4.8% (w/v) ethanol, 8.5% (v/v) phosphoric acid) was added to the BSA standards and protein samples in question. Both standards and samples were left at room temperature for 10 min before being read on the spectrophotometer (LambdaBio, Perkin-Elmer, Norwalk CT, USA) at 595nm. A standard graph of absorbance versus the amount of protein used was plotted. The amount of protein present from the unknown protein samples can then read from the standard graph.

2.2.3 Enzyme Assay for Pullulanase

A discontinuous assay modified from Bernfeld (1955) was used. 100µl of enzyme sample was incubated with 650µl of substrate (50mM Tris-HCl, pH 8.5, 6.7mM NaCl, 1% (w/v) Pullulan) at 80°C for 20 min. The reaction was stopped with 350µl of developing solution (1% (w/v) DNSA, 0.4M NaOH) and boiled at 100°C for 5 min. The assay samples were then cooled on ice before measuring the absorbance at 550nm. A blank was prepared by replacing the 100µl of enzyme sample with MQH_2O in the assay reaction.

A calibration curve was made by measuring differing amounts of maltose from 0-0.6mg using the same developing solution (Appendix II). One unit of enzyme activity is defined as 1µmole of reducing sugars released per min under assay conditions.

2.2.4 Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Buffers used in SDS PAGE are annotated in Table 2.7 and the reagents used in the making of polyacrylamide gel can be found in Table 2.8. A 10% running gel solution was prepared and poured between 2 glass plates sealed with a rubber seal. This gel was then overlaid with water-saturated butanol and left standing at room temperature for 20 min. The water-saturated butanol was then removed and the gel was washed with deionised water. A well-forming comb was then set up in between the glass plates and the stacking gel solution was added between the comb to form wells. Once the gel had polymerised, the gel with the glass plates was then loaded in the electrophoresis tank. The inner and outer reservoirs of the tank were filled with 1× tank buffer (Table 2.7) before the comb was removed. This mini gel apparatus is from ATTO Corporation, Nagoya, Japan.

10µl of sample containing 4-10µg of protein were mixed with 10µl of 2× loading buffer (Table 2.7) and heated at 100°C for 5 min prior to loading. Once samples were loaded, the gel was run at 10mA current (for single gel) at the stacking gel and 20mA once the protein reached the 10% running gel.

Once the electrophoresis was over, the gel was then gently removed from the plates and immersed in the coomasie staining solutions for 30 min. The gel was then washed in MQH_2O to remove residual coomasie stain before soaking in destain solution with gentle agitation. A small piece of sponge was also floated on the gel to soak up the coomasie stain that came off from the gel. After 1-1.5h incubation, the gel was then soaked in MQH_2O before framing in a membrane and dried overnight.

Table 2.7: Buffers used in SDS PAGE

Buffers	Ingredients	Special Notes
Running Gel Buffer	1.5M Tris, 0.4% (w/v) SDS	pH to 8.9 with HCl
Stacking Gel Buffer	0.48M Tris, 0.4% (w/v) SDS	pH to 6.8 with HCl
2× Loading Buffer	0.125M Tris, 4% (w/v) SDS, 20% (w/v) Sucrose, 0.08% (w/v) Bromophenol blue, 10% (w/v) β-Mercaptoethanol	pH to 6.8 with 1M HCl, then add in bromophenol blue and β-mercaptoethanol. Store in 4°C in 1ml aliquots
10× Tank Buffer	0.052M Tris, 0.1% (w/v) SDS, 0.4% (w.v) Glycine	To be used as 1× concentration
Coomasie Staining Solution	1.25g Coomasie Blue R, 227ml MQH_2O, 227ml Methanol, 46ml Acetic Acid	-
Destain Solution	75ml Acetic Acid, 50ml Methanol, 875ml MQH_2O	-

Table 2.8: Components used to make two 10% polyacrylamide gel.

Ingredients	Running Gel (10%)	Stacking Gel
Acrylamide (ml)	5	0.45
Gel Buffer (ml)	3.75	1.2
MQH_2O (ml)	6.25	1.8
10% (w/v) Ammonium persulphate (µl)	62.5	25
TEMED (µl)	15.6	5

Chapter 3: Growth of *Thermopallium natronophilum*

3.1 Introduction

In order to obtain native pullulanase and gDNA, growth of *Thermopallium* was established. *Thermopallium* is an obligate anaerobe and all growth conditions have to be carried out in serum bottles under oxygen free conditions.

3.2 Methods

3.2.1 Anaerobic Techniques and Equipment

Equipment used in the growth of *T. natronophilum* comprises four main parts. They are a gas proportioner, oxygen removal towel, a pressurised gassing manifold and gassing cannulas and a vacuum pump. The gas proportioner is used to control the mixture of H_2 and N_2 in the ratio of 4:1. A diagram of a gas proportioner is shown in Fig. 3.1 and the different gasses are supplied to the gas proportioner from the industrial tank (BOC Gas, UK). Trace oxygen is removed using oxygen removal tower, which is shown in Fig. 3.2. The oxygen removal tower consists of a copper column with copper-based catalytic pellets. Aluminium-coated glass wool is used to insulate the column. The reduction of the copper pellets is achieved by initially heating the pellets to a temperature of 150 °C under a flow of nitrogen at flow rate 0.5L/min. This is then followed by the introduction of hydrogen gas alongside nitrogen gas and the concentration of hydrogen is slowly increased. Complete reduction is achieved when nitrogen is completely displaced by hydrogen at 150 °C. The final part of the equipment consists of the pressurized gassing manifold and the gassing cannulas. The gassing manifold allows control of the introduction of different gasses in different ratios into serum bottles. A picture of the gassing manifold and

gassing cannulas is shown in Fig. 3.2. Disposable sterile needles can be attached at the ends of each gassing cannula and the needles can then be easily manipulated into the serum bottles though the rubber stoppers. For filtration of media into the closed top serum bottle though a 0.22μm filter (Milipore, UK), the vacuum pump can be turned on. This works on the concept that when the vacuum pump is turned on, air will be forced to move in the opposite direction, creating a low pressure within the serum bottle. By attaching a 50ml syringe with a 0.22μm filter, media with heat labile chemicals can then be filtered (Archaea, a Lab Manual: Methanogens, Cold Spring Harbor Laboratory Press).

3.2.2 Media Preparation

All growth conditions were performed in anaerobic conditions at 65°C, pH8.0 with continuous sparging with N_2 gas.

For a 100ml culture, 2.5g of starch, 0.2g of tryptone, 0.2g of NaCl, 0.1g of yeast extract and 250μl of resazurin at 0.2mg/ml were dissolved in 20ml of MQH_2O. This was then poured into a 120ml serum bottle (Supelco, UK), sealed with a rubber stopper and autoclaved. 0.5g Na_2CO_3, 250μl of solution A (Table 3.1), 50μl of solution B (Table 3.1), 50μl of solution C (Table 3.1), 25μl of vitamin solution (Table 3.1), 312μl of Cystein HCl (0.08g/ml), around 0.5ml of sodium sulphide (0.05g/ml) were dissolved in MQH_2O and made up to a final volume of 80ml. Following this, the 80ml growth media was filtered into the 150ml serum bottle via a 0.22μm filter, using the vacuum pump system described into the gassing manifold and gassing cannulas.

Before inoculation, growth media were continuously sparged with N_2 gas till the resazurin is completely reduced as indicated by a colourless state. For all growth, 10% inoculum was used.

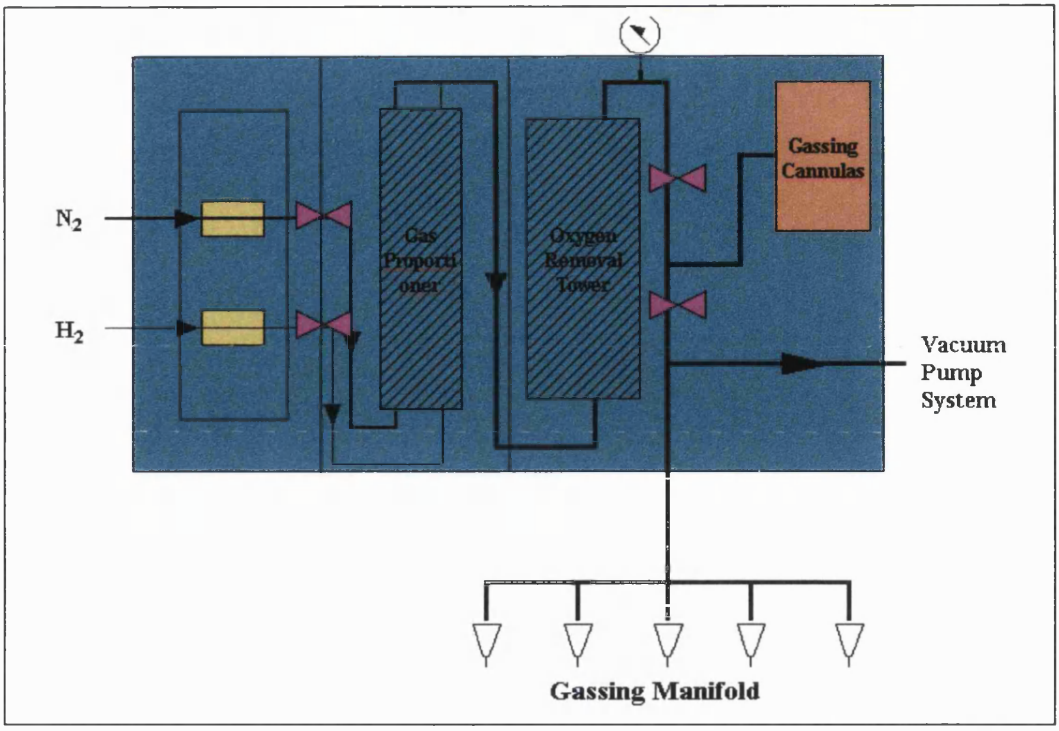


Fig. 3.1: Schematic representation of the anaerobic culturing equipments

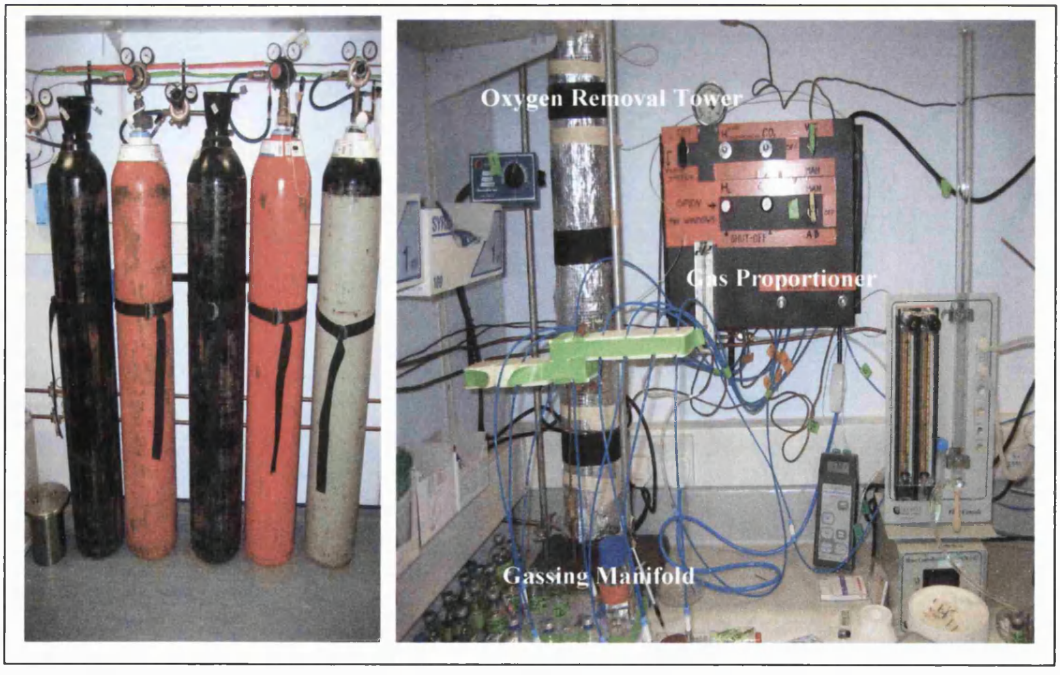


Fig. 3.2: Anaerobic equipment used for growth of *T.natronophilum*

Table 3.1: Components and instructions in making Vitamin Solution, Solution A, B and C; Sodium sulphide stock solution.

Solutions	Ingredients
Solution A (1L)	4g of trisodium citrate, 9g of $MnSO_4 \cdot 4H_2O$, 2.5g of $ZnSO_4 \cdot 7H_2O$, 2.5g of $NiCl_2 \cdot 6H_2O$, 0.3g of $AlK(SO_4)_2 \cdot 12H_2O$, 0.3g of $CoCl_2 \cdot 6H_2O$ and 0.15g of $CuSO_4 \cdot 5H_2O$ (Raven *et al.*, 1992).
Solution B (1L)	56g of $CaCl_2 \cdot 2H_2O$, 25g of NaBr, 16g of KCl, 10g of KI and 4g of $SrCl_2 \cdot 6H_2O$ (Raven *et al.*, 1992).
Solution C (1L)	50g of K_2HPO_4, 7.5g of H_3BO_3, 3.3g of $Na_2WO_4 \cdot 2H_2O$, 0.15g $Na_2MoO_4 \cdot 2H_2O$ and 0.005g of Na_2SeO_3 (Raven *et al.*, 1992).
Vitamin Solution (1L)	200mg of pyridoxine hydrochloride, 100mg of thiamine hydrochloride, 100mg of riboflavin, 100mg of nicotinic acid, 100mg of DL-calcium pantothenate, 100mg of lipoic acid, 40mg of biotin, 40mg of folic acid and 2mg of cyanocobalamin. . Riboflavin was first dissolved in 500ml of absolute ethanol whilst the rest were dissolved in water. The resulting solutions were then mixed together. All solutions were then filter-sterilised (Raven *et al.*, 1992).
Sodium Sulphide (0.05g/ml)	5g of sodium sulphide crystals were washed with MQH_2O and dabbed dry. A 100ml of MQH_2 was then boiled to remove dissolved oxygen. Sodium sulphide stock is then kept in an air-tight serum bottle and flushed with N_2 (g).

3.2.3 Small-scale growth of *T. natronophilum*

Small-scale growth involves only 50ml or 100ml of growth medium. No continuous sparging is required. Small-scale growth is normally carried out in order to produce an inoculum for subsequent large-scale production. Cells are ready to be used as inoculum after 17 h incubation at 65°C.

3.2.4 Large-scale growth of *T. natronophilum*

Large-scale growth of *T. natronophilum* is carried out in 1L serum bottle (OCHS, Germany). The culture is continuously sparged with N_2 gas. An outlet, which involves a needle attached to tubing, is also introduced in each individual serum bottle. A cool water jacket surrounds the outlet tubing and this is to reduce evaporation of the growth media. All bottles are immersed in a water bath at 65°C.

3.2.5 Obtaining growth curve for *T. natronophilum*

For obtaining a growth curve of *T.natronophilum*, a 400ml sample of medium was prepared. 10% inoculum was used and the culture was continuously sparged with N_2. An initial OD reading was taken and a 1ml sample was then taken at the appropriate time points till the stationary phase of the cell growth was achieved.

3.2.6 OD measurement of *T. natronophilum* cells

For OD measurement, samples have to be taken out from the serum bottle without disturbing the anaerobic conditions in the bottle. Anaerobic conditions were maintained by over pressurising the bottle with N_2 gas and a syringe was then introduced into the bottle. Due to the over pressure inside the serum bottle, cell culture will automatically be transferred into the syringe if the needle is in the media. Approximately 2ml of growth culture is needed. From here, 1ml of the cell culture is spun down and 500µl of the supernatant is then mixed with 500µl of MQ H_2O and used as blank. A few granules of sodium dithionite, $Na_2S_2O_4 \cdot H_2O$ are then added. Sodium dithionite is a strong reducer and should turn the resazurin in the blank from

pink to a colourless solution. For measuring OD of samples, 500µl of cell culture is mixed with 500µl of MQ H_2O. $Na_2S_2O_4 \cdot H_2O$ is also added to the samples. OD of all samples was measured at absorbance 600nm.

3.2.7 Inspection of cells under microscope

A drop of cell culture was put on a slide and inspected under a Nikon Eclipse E200 light microscope (Nikon, Japan) under 400× magnification. This is to check the stress level of *T.natronophilum*. *T.natronophilum* will change its morphology from rod shaped cells to cocci shaped cells under stressful conditions or when stationary phase is reached.

3.3 Results

3.3.1 Optimising the Growth Conditions of *T.natronophilum*

During media preparation, several problems were encountered. In the first attempt, 5ml of solution A, 0.5ml of solution B and 0.5ml of solution C gave rise to a white precipitate. In addition, a black precipitate was observed when 1ml of sodium sulphide (0.25g/ml) was added to the mixture. In order to get the optimum media conditions, different volumes of each solutions used were investigated. The following table shows the variation used.

Table 3.2: The different materials and the different ratios of material used in the growth of *T.natronophilum*

	Combination 1	**Combination 2**	**Combination 3**
Solution A	50µl	250µl	250µl
Solution B	50µl	50µl	50µl
Solution C	50µl	50µl	50µl
Vitamin Solution	25µl	25µl	25µl
Sodium Sulphide(0.05g/ml)	500µl	500µl	250µl
Cysteine-HCl (0.08g/ml)	-	-	160µl

In all 3 combinations, no white precipitate was observed but black precipitates were found in solution with combination 1 and 2. From here, the ratio of the different solutions in combination 3 was then used for subsequent growth of *T.natronophilum*. In conclusion, combination 3 gives the best growth conditions for *T.natronophilum*.

Initially all small or large-scale growth of *T.natronophilum* was done in a closed system. The cells normally reached an OD_{600} of 0.3-0.4 and when observed under the

light microscope, cells were no longer dividing and took coccoid morphology (stress or stationary phase). This is typical of all *Thermotogales* (Huber *et al.*, 1986). To overcome this limitation, continuous sparging of N_2 gas into the serum bottles was employed. With this improvement, cells were growing to an OD_{600} of 0.6.

3.3.2 Growth Profile of *T.natronophilum*

The growth profile of *T.natronophilum* was investigated (Fig. 3.3).

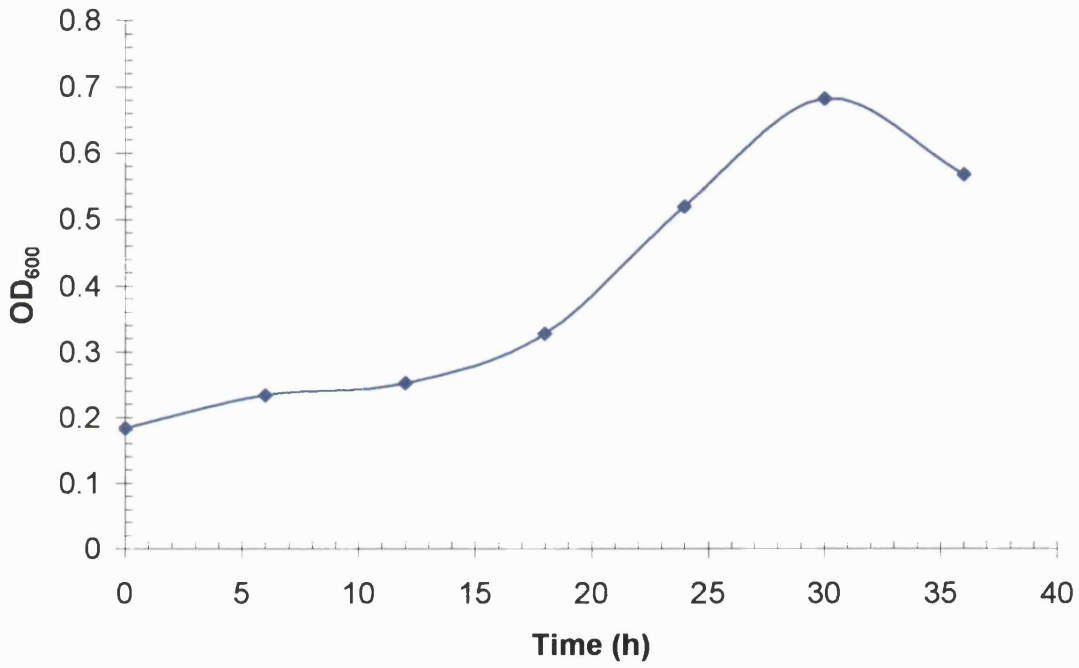


Fig. 3.3: Growth curve of *T.natronophilum*

T.natronophilum has a lag phase of 12 h, and a log phase of 18 h. Cells are typically harvested after 28 h of growth, which is 2 h before the stationary phase. Based on microscopic observation, cells will get stressed during stationary phase and no longer take the morphology of a rod shaped cell.

Chapter 4: Locating the Pullulanase Type I Gene of *T. natronophilum*

4.1 Introduction

From previous work done by Dr. Carl Thompson, the native pullulanase of *T. natronophilum* has been purified and characterised. Furthermore, a 20 residue internal amino acid sequence of the pullulanase (YIGDGAWEAVLEGDDEGXFYR) has been successfully obtained (Thompson, 1998). Considering the potential biotechnological application of this enzyme, it is important to obtain the recombinant form of this enzyme for it to be characterised and later mass-produced. To do this, the gene encoding pullulanase has first to be obtained, cloned and expressed in *E.coli*. Following this, characterisation of the recombinant protein is to be carried out and a comparison of the recombinant and native enzyme can be realised.

This chapter is subdivided into four sections with each section showing its own materials and method, followed by results supporting the methods used. The first section involves the amplification of a probe for the pullulanase gene using PCR. This was done by amplifying part of the gene from *Thermopallium* gDNA, using degenerate primers designed from the internal amino acid sequence and homologous region of type I pullulanase from other strains. This attempt was successful and the probe was then fluorescently labelled using the DIG system. The second section of this chapter annotates the making of a gDNA library and the probing of the library with DIG-labelled pullulanase probes (Höltke *et al.*, 1995). The screening of the gDNA library with the probe was not successful, and so the same library was screened with nested primers in colony PCR. The latter approach also did not yield any positive results.

The third section depicts how the complete gene sequence of pullulanase was obtained by PCR, using degenerate and specific primers. The majority of the gene sequence towards the 5' end and 3' end was successfully obtained using this method.

The fourth section investigates how the start and stop codons of the gene were obtained. Inverse PCR and direct sequencing of the gDNA with specific primers were attempted but to no avail. Finally, the approach that yielded results involved the use of biotinylated primers and non-specific primers in PCR. Subsequently, the selection of positive PCR products was carried out using streptavidin coated paramagnetic beads. Specific primers were then used to sequence into the unknown regions of gene sequence.

Section 1: Amplification and DIG Labelling of Pullulanase Gene Probe

4.2 Methods

4.2.1 Obtaining a Probe for the Pullulanase Gene with PCR

PCR was carried out as described in section 2.2.6. gDNA from *Thermopallium* was used as template DNA. Degenerate primers used for the amplification are illustrated in Fig.4.1. Single primer controls were also set up for the PCR reaction. PCR was carried out under 1 cycle of 96°C for 3 min, 30 cycles of 96°C for 1 min 30 sec, 55°C for 1 min, 72°C for 1 min and finally 1 cycle of 72°C for 10 min. Then, 10µl of the reaction was run on 1% agarose gel.

Internal Amino Acid: Pullulanase Forward primer 1

```
      Y   I   G   D   G   A   W
5'-TAC ATH GGW GAY GGW GCW TGG-3'
     Degeneracy: 48 Tm: 57.6°C
```

Homologous Region: Pullulanase Reverse Primer 1 (Reverse and Complementary)

```
       M   M   R   K   Y   I   V   D   T   L
(N) 5'-ATG ATG AGR AAA TAC ATM GTT GAY ACA CTY-3'
(C) 3'-TAC TAC TCY TTT ATG TAK CAA CTR TGT GAR-5'
(R) 5'-RAG TGT RTC AAC KAT GTA TTT YCT CAT CAT-3'
          Degeneracy: 16 Tm: 58.5°C
```

Fig. 4.1: Degenerate primers used to obtain the probe. In the reverse primer, (N) annotates the sequence from the positive strand from 5' to 3'. (C) annotates the complementary sequence of (N), and (R) is the reverse sequence of (C). The single letter codes used in degenerate primers can be found under abbreviations on page vi.

4.2.2 DIG Labelling the Pullulanase Gene Probe

1μg of pullulanase gene probe was prepared by PCR amplification as described in section 4.2.1. The gene probe was then random primed labelled using DIG HIGH PRIME DNA Labelling and Detection Started Kit II (Roche, Mannheim, Germany), following the manufacturer's instruction. The labelling reaction was incubated for 16h whereby the probe was labelled with digoxigenin-11-dUTP by Klenow enzyme.

4.2.3 Quantification of Labelling Efficiency and Detection Efficiency

The yield of labelled probed was quantified as described by the manufacturer's instructions whereby the DIG labelled control DNA and DIG labelled pullulanase gene probe were diluted to 1ng/μl, 10pg/μl, 3pg/μl, 1pg/μl, 0.3pg/μl, 0.1p/gμl, 0.03pg/μl and 0.01pg/μl. 1μl of each diluted sample was then blotted onto the positively charged nylon membranes (Hybond^{TM}-N+, Amersham Pharmacia Biotech, UK) and the nucleic acids were fixed to the membrane via UV cross linking at 1200J. The DIG labelled control DNA was used as a standard to determine the true amount of pullulanase gene probe that had been successfully labelled by comparing the intensities of each dot generated.

The sensitivity of the labelled probe in detecting the unlabelled pullulanase probe was also determined. The same series of concentrations of unlabelled probes was used. The same DIG labelled DNA probes were used as the standard of comparison.

4.3 Results

4.3.1 Obtaining Pullulanase Gene Probe

In order to locate the gene sequence for pullulanase type I of *Thermopallium natronophilum*, a probe for the gene has to be obtained. This probe was successfully obtained via PCR. To start with, the 20 internal amino acid of pullulanase (YIGDGAWEAVLEGDDEGXFYR) was aligned to the pullulanase type I protein sequence of *F. pennavorans Ven*5. The 20 amino acids were found to align to amino acid 277-283 of pullulanase of *F. pennavorans Ven*5 protein sequence. YIGDGAW was then chosen as the basis for the degenerate forward primer for the first round of PCR (Blue box in Fig.4.2). The degenerate reverse primer was designed from the homologous region MMRKYIVDTL from the alignment of pullulanase type I of *F. pennavorans Ven*5 and *T.maritima* (Red box in Fig. 4.2).

From the pullulanase type I alignment in Fig. 4.2, the degenerate primers should produce a PCR product of approximately 760bp. The single band of approximately 700bp shown in Fig. 4.3 corresponds to the predicted size and was thus excised, gel purified, and was sequenced with the same degenerate primers.

The PCR product yielded a 720bp nucleotide sequence, which was then translated to protein sequence and aligned with the pullulanase type I protein sequences of *F. pennavorans Ven*5 and *T.maritima,* (Fig. 4.4) using GCG programme Pile-Up (Genetics Computer Group, Inc, Madison, Wisconsin, USA). The output file in Fig. 4.4 is presented in GeneDoc Version 2.6.001 (Nicholas *et al.*, 1997) format. The protein sequence obtained was found to have 75% sequence identity with pullulanase type I of *F.pennavorans* and 65% sequence identity with pullulanase type I of *T.maritima*. This suggested that the amplified product was indeed part of the pullulanase type I gene.

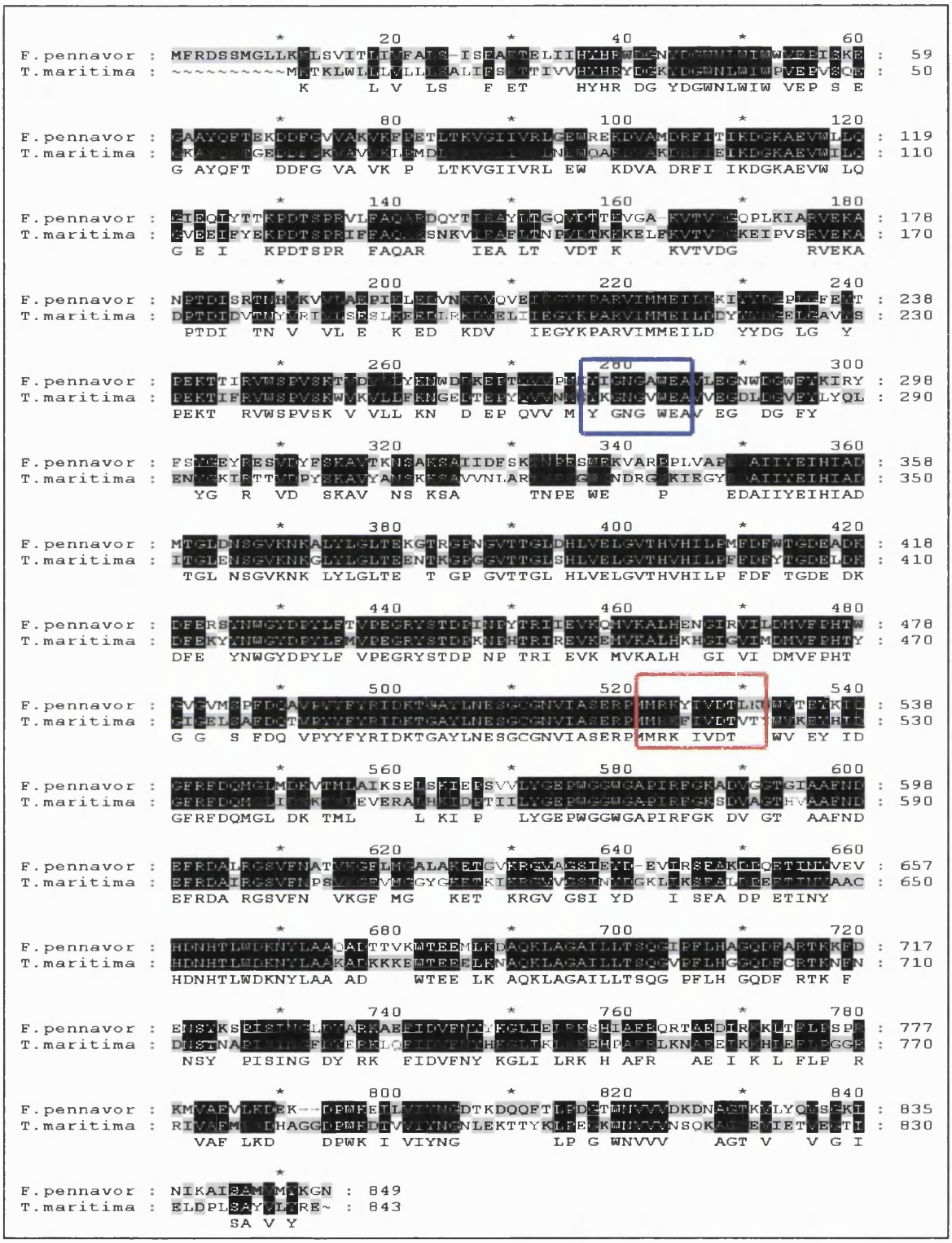


Fig. 4.2: Alignment of the complete pullulanase protein sequence of *Fervidobacterium pennavorans Ven*5 and *Thermotoga maritima*. The blue box identifies where the *Thermopallium* internal pullulanase peptide sequence (YIGDGAW) is homologous with the two other strains. The red box identifies the peptide sequence that was used to design degenerate reverse primer to obtain the pullulanase probe.

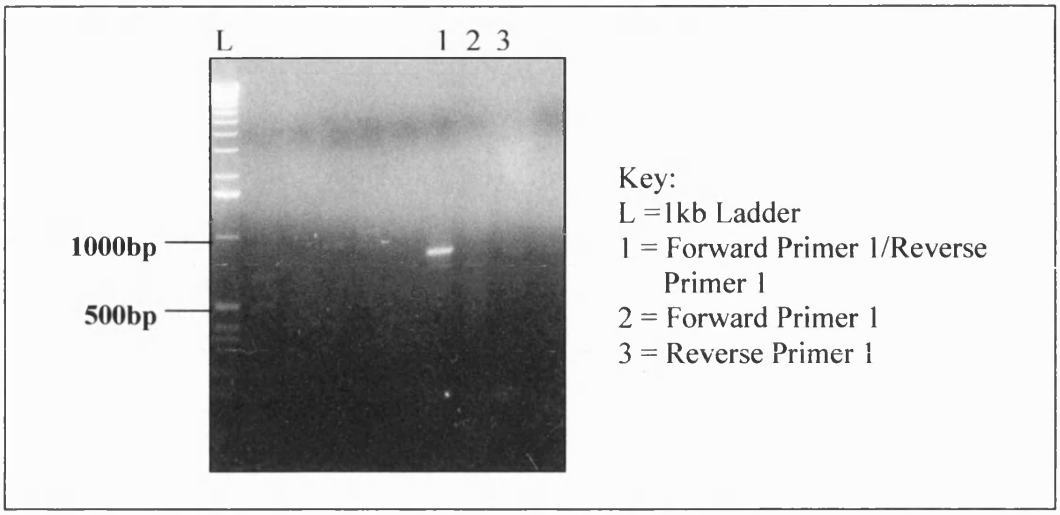


Fig 4.3: Agarose gel electrophoresis of the PCR product obtained from the degenerate primers designed in Fig. 4.1. Lane 1 is the sample from the double primer PCR reaction. Lane 2 and Lane 3 are samples from single primer controls.

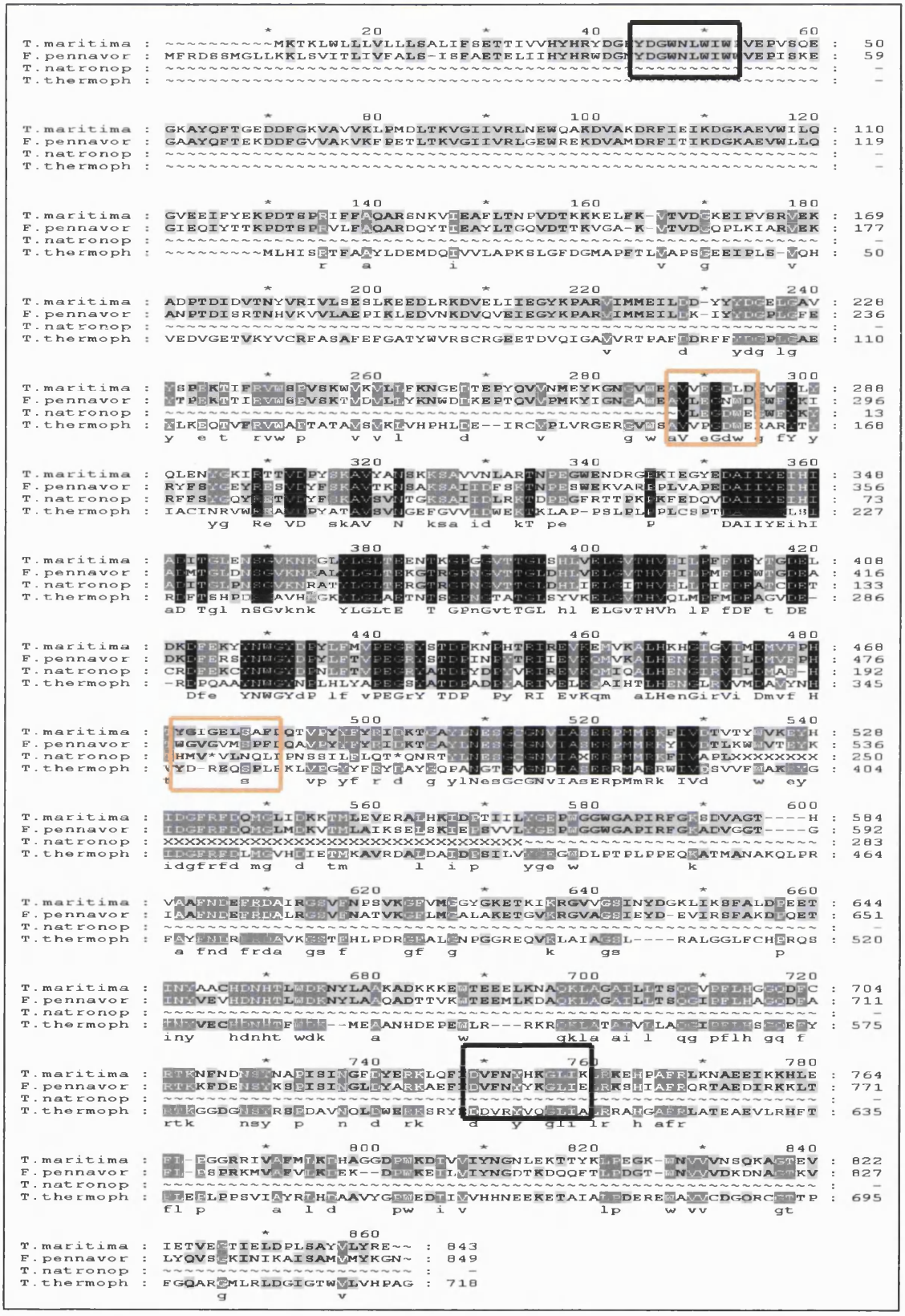


Fig. 4.4: PCR product (pullulanase gene probe sequence) obtained was sequenced and aligned with pullulanase type I of *T.maritima*, *F.pennavorans* and *T.thermophilus*. Only strict identities were shaded black. The highlights decrease in intensity when there is less homology in amino acids. Sequences boxed in yellow were used to design specific primers while the sequences boxed in black were used to design degenerate primers for PCR.

4.3.2 DIG-Labelling the Pullulanase Gene Probe

Following the successful amplification of the pullulanase gene probe, 1µg of the probe was prepared. Subsequently, the probe was labelled using the DIG system. Initially, the efficiency of the DIG labelling of pullulanase gene probe was investigated. It was found that approximately 10pg/µl of pullulanase gene probe was labelled (Fig. 4.5A) and the labelled probe was able to detect down to 10pg/µl of unlabelled probe (Fig. 4.5B).

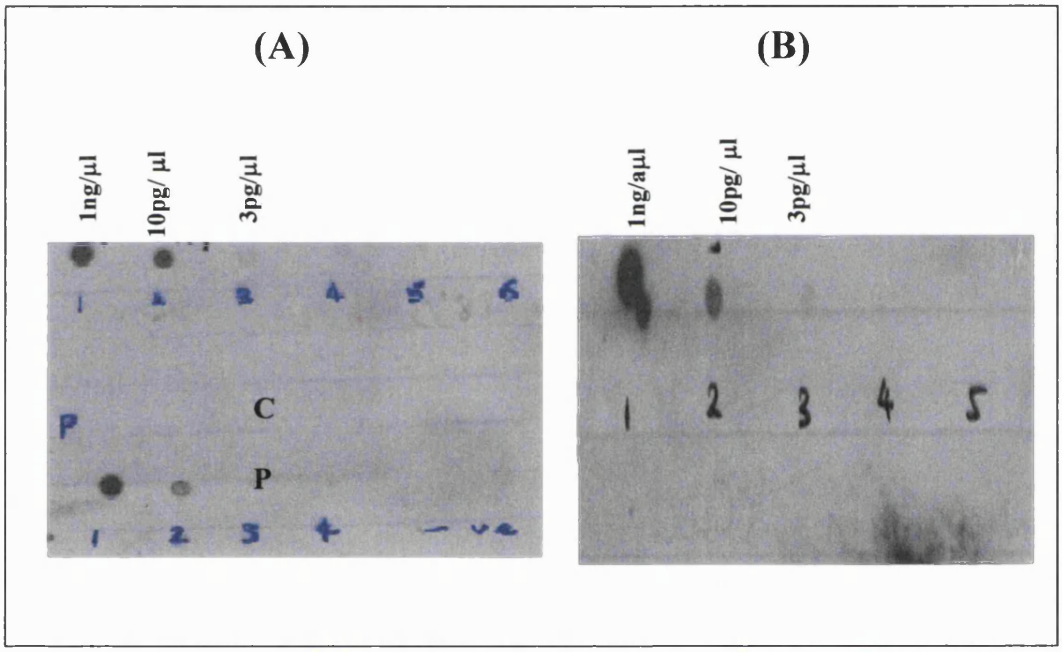


Fig. 4.5: (A) Determination of the labelling efficiencies of the probe with DIG High Probe DNA Labelling and Detection Starter Kit II. C represents the control and P represents the labelled pullulanase probe.

(B): Determination of the efficiency of the probe detecting the unlabeled probe.

Section 2:The Making and Screening of Genomic DNA Library

4.4 Methods

4.4.1 Restriction Digestion of gDNA with *Msp*I to Produce 10kb Fragments

gDNA extraction of *Thermopallium* was carried out according to the method described in section 2.2.2. 100µl of gDNA (77µg) was mixed with 15µl of 10× NEB Buffer 2, 35µl of MQH_2O and heated to 72°C for 10min to denature the gDNA. 130µl of the mixture was aliquoted into 13 tubes with 10µl in each tube and kept on ice. The remaining 20µl was mixed with 10U of *Msp*I. 10µl of this mixture was then transferred to the first aliquoted tube. The content in the first tube was mixed and 10µl of it was then transferred to the second aliquoted tube. The content in the second tube was then mixed and 10µl of this was transferred to the third aliquoted tube. This was repeated in all subsequent tubes and incidentally all the enzymes were serially diluted (2× dilution) in each stage. All tubes were then incubated at 37°C for 30min, followed by incubation at 65°C for 20min to inactivate the enzyme. 10µl of each sample was then loaded onto a 0.7% agarose gel. DNA digested to about 10kb in length was then excised and gel purified using Qiaex II DNA Extraction Kit (Qiagen, Germany).

4.4.2 Digestion and Dephosphorylation of pTrueBlue-*rop* Plasmid for gDNA Library Generation

The pTrueBlue-*rop* plasmid (Genomics One Corporation, Quebec, Canada) is a low copy number plasmid derived from pBR322 and offers the user stringent blue/white selection in DNA cloning.

10U of *Cla*I was used to digest 1µg of pTrueBlue-*rop* plasmid. Digestion was set up as described in section 2.2.5. The reaction was incubated at 37°C for 1h and subsequently the whole mixture was loaded and run on 0.7% agarose gel. Linearised pTrueBlue-*rop*, which gives a fragment size of 4kb, was excised and gel purified using Qiaex II DNA Extraction Kit and resuspended in MQH_2O. Then, the linearised

plasmid was dephosphorylated (section 2.2.9) with shrimp alkaline phosphatase (Roche, Mannheim, Germany) at 37°C for 1h. The enzymatic reaction was stopped by heating the mixture to 65°C for 20min. Following this, the plasmid was gel purified and quantified by running a small sample on a 0.7% agarose gel alongside a DNA mass ladder.

4.4.3 Ligation of 10kb gDNA fragments into pTrueBlue-*rop* plasmid

Firstly, the concentration of insert DNA and plasmid DNA were quantified (section 2.2.3). Ligation was carried out at vector: insert ratios of 1:3, 1:2 and 1:1. Determination of each ratio was calculated using the formula described in section 2.2.11. The ligase used is Epicentre Fast-Link ™ DNA Ligation Kit from Cambio, Cambridge, UK. The kit is supplied with its own 10× reaction buffer (to be diluted down to 1× final concentration) and 10mM ATP (to be diluted to1mM final concentration). Typically, a ligation reaction was in a final volume of 15µl and was incubated at room temperature for 30 min and enzymatic inactivation was carried out at 70°C for 15 min. For each ligation, 2 controls were set up. For the negative control, unphosphorylated linearised pTrueBlue-*rop* was used, and for the positive control phosphorylated, linearised pTrueBlue-*rop* was used. These ligation reactions were then transformed *into E.coli* XL1-Blue cells by electroporation.

4.4.4 Transformation of *E.coli* XL1-Blue Cells by electroporation.

All 1.5ml tubes and electroporation cuvettes (Bio-Rad Laborotaries GmbH, München, Germany) were chilled while the electrocompetant cells were thawed on ice. For transformation, $1/10^{th}$ of the ligation reaction was mixed with 40µl of *Epicurion coli*® electrocompetant XL1-Blue cells in chilled 1.5ml tubes before transferring them to the cuvette chamber prior to electroporation. The control for transformation was 10ng of uncut pUC18. A MicroPulser™ (Bio-Rad Laborotaries GmbH, München, Germany) was employed and transformation was carried out with a pulse of 0.75ms at 1.5KV. After pulsing, 960µl of 37°C pre-warmed SOC medium was added to the

cuvette chamber and 100µl of this was then plated onto LB-Amp plates, with 0.003%(w/v) X-Gal and 1mM IPTG and incubated overnight at 37°C.

To facilitate the screening of the library, all white colonies were picked and transferred into 96 well plates supplied with 200µl of LB-Amp. Each plate was then covered with a Breath-Easy Sealing Membrane (Diversified Biotech, Sigma, UK) and grown overnight at 37°C under shaking at 325rpm.

4.4.5 Screening of the Genomic Library

4.4.5.1 Screening of the Genomic Library with DIG-Labelled Pullulanase Gene Probe

Firstly, positively charged nylon membranes (Hybond^{TM}-N+, Amersham Pharmacia Biotech, UK) were laid over rectangular LB-Amp plates. The cells in 96 well plates were transferred onto the membranes with a 96-pin replicator (Boekel Scientific, Feasterville, PA, USA) and incubated at 37°C till visible colonies were formed. The colonies formed on the membrane were lysed by soaking the membranes in a few solutions (stated below) sequentially. At each stage, excess fluid was removed by sliding the filter with forceps over the edge of the soaking tray. The solutions are 1.2ml of 0.5M NaOH for 7 min, 1.5ml of 0.1M NaOH for 5 min, 5ml of 0.2M Tris-HCl, pH 7.5, 2mM EDTA for 3 min and finally 5ml of 2× SSC (17.53g/L NaCl, 8.82g/L NaCitrate, pH 7.0) containing 25mM of sodium phosphate, pH 7.2, and 2mM EDTA for 3 min. The membranes were then left to dry on filter paper for 10 min. For the positive control, 1µl (10ng) of the unlabelled probe was then dotted on the left hand upper corner of each membrane. The DNA was cross-linked to the membranes by exposing each membrane to 1200J of UV light. These membranes were then stored at 4°C in plastic films till needed.

Prior to screening of the membrane with the labelled probe, the membrane was washed twice in 20ml of 2× SSC, 0.1% SDS, at room temperature for 5 min under constant agitation. This was then followed by two washes of the membrane with

0.5% SSC, 0.1% SDS and 65°C for a 15min period under constant agitation. For immunological detection with labelled probe, the membrane was rinsed in washing buffer (0.1M maleic acid, 0.15M NaCl, pH 7.5, 0.3%(w/v) Tween20) for 5 min, 100ml of 1× Blocking Solution (supplied by manufacturer as a 10× stock which was diluted down with maleic acid buffer [0.1M maleic acid, 0.15M NaCl, pH adjusted to 7.5 with solid NaOH]) for 30 min, 10ml of antibody solution (supplied by manufacturer to be made up to a concentration of 75mU/ml in blocking solution prior to use) for 30 min, two washes with 100ml of washing buffer for 15 min each, and finally 20ml of detection buffer (0.1M Reis-HCl, 0.1M NaCl, pH 9.5) for 5 min. All the washings in immunological detection were done at 45°C. All washings were done in hybridisation tubes (Fisher, UK) and were incubated at the required temperature in the OV4 Compact Line Hybridisation Oven (Biometra GmbH I. L., Goettingen, Germany). After the serial washing, the membrane was placed between a plastic folder and the membrane was then covered with CSPD solution supplied by the manufacturer. Excess fluids and air bubbles were gently pressed out of the folder. The membrane was then incubated at 37°C for 10 min before exposing the membrane to Hyperfilm-ECL (Amersham, Sweden) for 1.5 h.

4.4.5.2 Screening of the Genomic Library Using Colony PCR

This was carried out as described in section 2.2.6.6. 20µl from each well in each row were pooled together and boiled for 10min and cell debris was spun down. 5µl of the supernatant was then used as template in PCR with pullulanase forward and reverse primer 1 (Fig. 4.1).

4.5 Results

4.5.1 Making the Library and Library Screening

After establishing the DIG-labelled pullulanase gene probe, a gDNA library that consisted of plasmid vectors with 10-15kb inserts was made. A pilot restriction digest of the gDNA with *Msp*I was carried out and 0.01U, 0.02U and 0.05U were found to give the best results (data not shown). Following establishing the optimum conditions for gDNA digestion, the reactions were scaled up 10× to generate a stock of 10kb gDNA. The digested gDNA was then used to ligate into pTrueBlue-*rop* vector that was digested with *Cla*I and dephosphorylated with shrimp alkaline phosphatase. A vector: insert ratio of 1:3 gave the biggest number of positive clones. Both blue and white colonies were observed in LB-Amp plates. All white colonies were picked into 96-well plates. A total of 18 96-plates were made and were screened with DIG-labelled pullulanase gene probe. No positive colonies were picked up using this approach. The positive control (unlabelled probe) dotted to the membrane was detected by the DIG-labelled pullulanase gene probe. To double-check the library for positive clones, colony PCR was carried out. 8 clones in each row were pooled together and the supernatant was used as template for PCR. Thus in each PCR reaction, a total of 8 clones will be tested. The positive control for colony PCR was done using gDNA as template. From this experiment, no positive clones were located in the gDNA library. The positive control showed that the PCR reaction itself was successful.

Section 3: Amplification of the Complete Pullulanase Gene Sequence Using PCR

Due to time constraints, another approach was used to locate the rest of the pullulanase gene. The majority of the pullulanase gene was successfully obtained via PCR amplification. This was made possible by designing specific primers from the probe sequence whilst the other primers were designed from homologous region of the pullulanase type I gene from *T.maritima*, *F.pennavorans* and *T.thermophilus*.

4.6 Methods

4.6.1 Locating Pullulanase Gene Sequence with PCR Using Sets of Degenerate and Specific Primers

PCR was carried out as described in section 2.2.6. Primers used are shown in Fig. 4.6. Primers designed from homologous regions were degenerate and the preferred codons of *Thermopallium* were used. A statistical calculation of the preferred codon usage of *Thermopallium* was carried out based on the gene sequence of the pullulanase probe (Table 4.1). In Fig. 4.5, peptide sequences boxed in orange were used to design specific primers, while peptide sequences boxed in black were used to design degenerate primers. The primers used for amplification of the pullulanase gene towards the 5'end were designed from GWNLWIW and VLEGDWEG.

For obtaining the pullulanase gene sequence towards the 5' end, PCR was carried out under 1 cycle of 96°C for 3 min, 30 cycles of 96°C for 1 min 30 s, 59.1°C for 1 min, 72°C for 1 min and finally 1 cycle of 72°C for 10 min. For obtaining the pullulanase gene sequence towards the 3' end, annealing was carried out at 59.4°C for 1min 15 s instead.

Towards 5' End:

Homologous Region Pullulanase Forward Primer 2 (Degenerate)

```
      G   W   N   L   W   I   W
5'-GGT TGG AAC CTY TGG ATH TGG -3'
  Degeneracy: 6        Tm: 59.6°C
```

Sequence from Gene Probe: Pullulanase Reverse Primer 2 (Specific) (Reverse and Complementary)

```
        V   L   E   G   D   W   E   G
(N) 5'-GTT TTG GAA GGT GAT TGG GAA GG—3'
(C) 3'-CAA AAC CTT CCA CTA ACC CTT CC-5'
(R) 5'-CC TTC CCA ATC ACC TTC CAA AAC-3'
     Degeneracy: 1        Tm: 62.6°C
```

Towards 3' End:

Sequence from Gene Probe: Pullulanase Forward Primer 3 (Specific)

```
      R   K   F   I   V   A   P
5'-TGA GAA AAT TCA TTG TGG CAC CC -3'
  Degeneracy: 1        Tm: 62.4°C
```

Homologous Region Pullulanase Reverse Primer 3(Degenerate) (Reverse and Complementary)

```
        I   D   V   F   N   Y   Y   K   G   L
(N) 5'-ATH GAY GTT TTC AAC TAC TAC AAA GGW CTT-3'
(C) 3'-TAD CTR CAA AAG TTG ATG ATG TTT CCW GTT-5'
(R) 5'-TTG WCC TTT GTA GTA GTT GAA AAC RTC DAT-3'

          Degeneracy: 24       Tm: 61.3°C
```

Fig. 4.6: Primers used to obtain the 5'end and the 3' end of the pullulanase sequence, (N) annotates the sequence from the positive strand from 5' to 3'. (C) annotates the complementary sequence of (N), and (R) is the reverse sequence of (N). The single letter codes used in degenerate primers can be found under abbreviations on page vi.

Table 4.1: Codon frequency calculated from pullulanase gene probe. These values are obtained from GCG. Column number 3 depicts the number of each amino acid present in the sequence, and fraction is the value generated when the number of a particular amino acid is divided by the total number of the same amino acid, including the degenerate codons.

Amino Acid	Codon	No.	Fraction
Gly	GGG	1	0.06
Gly	GGA	8	0.44
Gly	GGT	8	0.44
Gly	GGC	1	0.06
Glu	GAA	15	1.00
Asp	GAT	11	0.73
Asp	GAC	4	0.27
Val	GTG	3	0.19
Val	GTA	3	0.19
Val	GTT	7	0.44
Val	GTC	3	0.19
Ala	GCG	3	0.27
Ala	GCA	6	0.55
Ala	GCT	2	0.18
Arg	AGG	3	0.20
Arg	AGA	12	0.80
Ser	AGT	3	0.38
Ser	AGC	1	0.12
Lys	AAG	3	0.30
Lys	AAA	7	0.70
Asn	AAT	7	0.58
Asn	ATC	5	0.42
Met	ATG	5	1.00
Ile	ATA	7	0.41
Ile	ATT	8	0.47
Ile	ATC	2	0.12

Amino Acid	Codon	No.	Fraction
Th	ACG	1	0.05
Th	ACA	6	0.32
Th	ACT	8	0.42
Th	ACC	4	0.21
Trp	TGG	3	1.00
Cys	TGT	3	0.75
Cys	TGC	1	0.25
Tyr	TAT	5	0.38
Tyr	TAC	8	0.62
Leu	TTG	2	0.10
Leu	TTA	6	0.30
Phe	TTT	6	0.46
Phe	TTC	7	0.54
Ser	TCT	3	0.38
Ser	TCC	1	0.12
Gln	CAG	3	0.43
Gln	CAA	4	0.57
His	CAT	2	0.29
His	CAC	5	0.71
Leu	CTA	2	0.10
Leu	CTT	6	0.30
Leu	CTC	4	0.20
Pro	CCG	4	0.31
Pro	CCA	5	0.38
Pro	CCT	3	0.23
Pro	CCC	1	0.08

4.7 Results

4.7.1 PCR Amplification of the Pullulanase Gene Towards the 5' End

Based on the alignment in Fig. 4.4, the predicted size obtainable from primers GWNLWIW (degenerate) and VLEGDWEG (specific) will be 747bp with respect to the protein sequence of *F.pennavorans Ven*5. From Fig. 4.7, the band between 700bp and 800bp corresponds to the predicted size and no bands of similar size were found in the single primer controls in lane 2 and 3. This PCR product was then excised from the gel, purified and sequenced. From the sequencing data obtained, the nucleotide sequence was translated to protein sequence and aligned with pullulanase type I protein sequence from *F.pennavorans Ven*5. From the alignment obtained (Fig. 4.8), the amplified product is indeed part of the pullulanase type I family gene sequence and aligns with amino acids 56 to 253 of the *F.pennavorans Ven*5 enzyme.

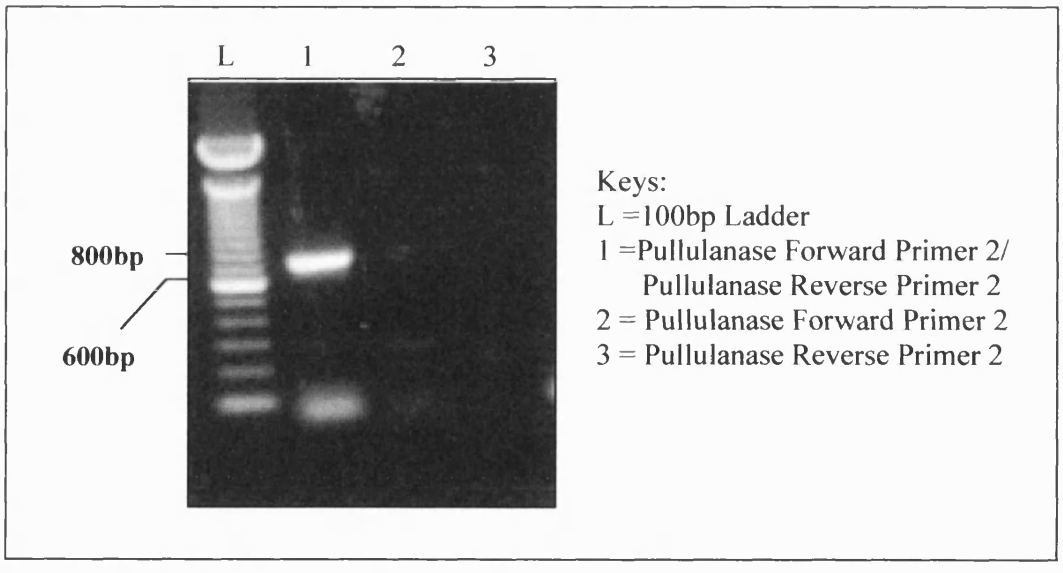


Fig. 4.7: Amplification of the pullulanase gene towards the 5'end using primers Pullulanase Forward Primer 2 and Pullulanase Reverse Primer 2. Refer to Fig. 4.6 for the primers used.

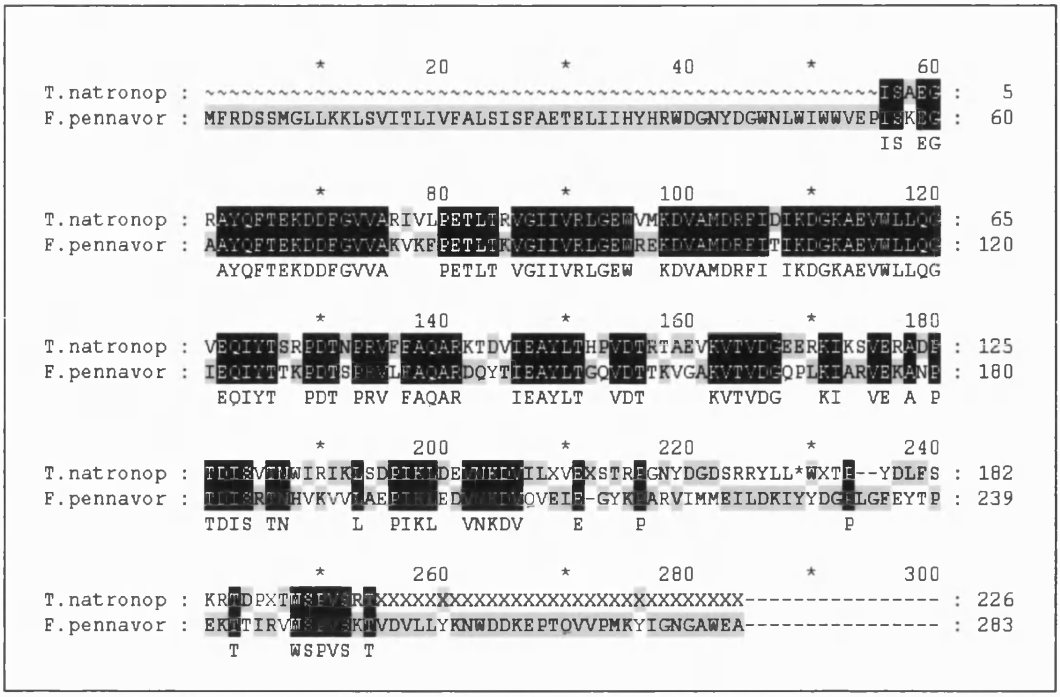


Fig. 4.8: The protein sequence alignment of PCR product amplified in Fig. 4.7 with pullulanase protein sequence of *F.pennavorans Ven*5.

4.7.2 PCR Amplification of the Pullulanase Gene Towards the 3' End with PCR

To amplify the gene sequence towards the 3'end, the amino acid sequences RKFIVAP (specific) and IDVFNYYKGL (degenerate) (Fig. 4,6) were used to design PCR primers. Based on the alignment in Fig. 4.4, the predicted size obtainable from this pair of primers will be 828bp with respect to the gene sequence of *F.pennavorans Ven*5. Using both primers a PCR product of approximately 850bp was successfully amplified (Fig. 4.9), which is in good agreement with the predicted size. An alignment of the protein sequence derived from this PCR product and the protein sequence of pullulanase type I of *F.pennavorans Ven*5 has confirmed the results obtained (Fig. 4.10). The PCR product amplified aligns with amino acids 488 to 739 of the *F.pennavorans Ven*5 enzyme.

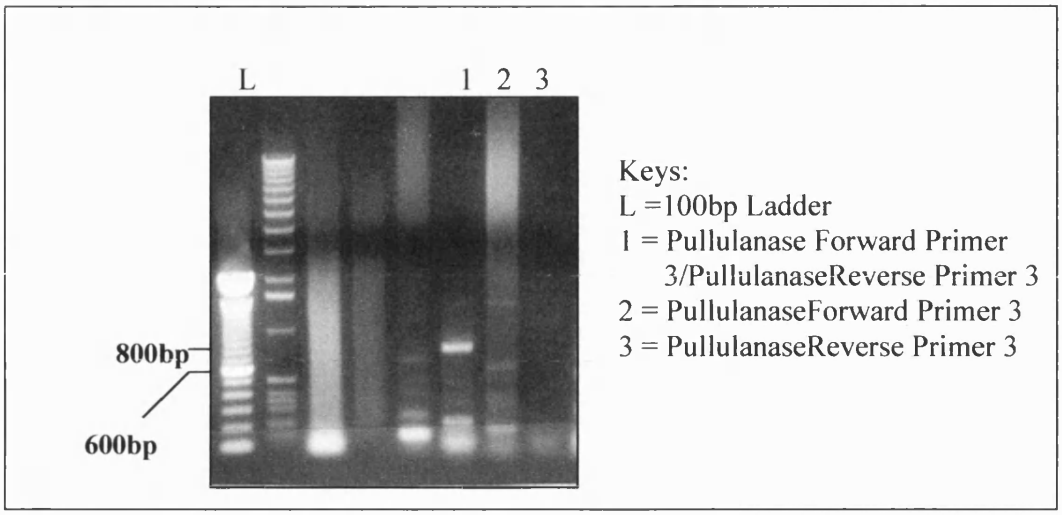


Fig. 4.9: Amplification of the pullulanase gene towards the 3'end using primers Pullulanase Forward Primer 3 and Pullulanase Reverse Primer 3. Refer Fig. 4.6 for the primers used.

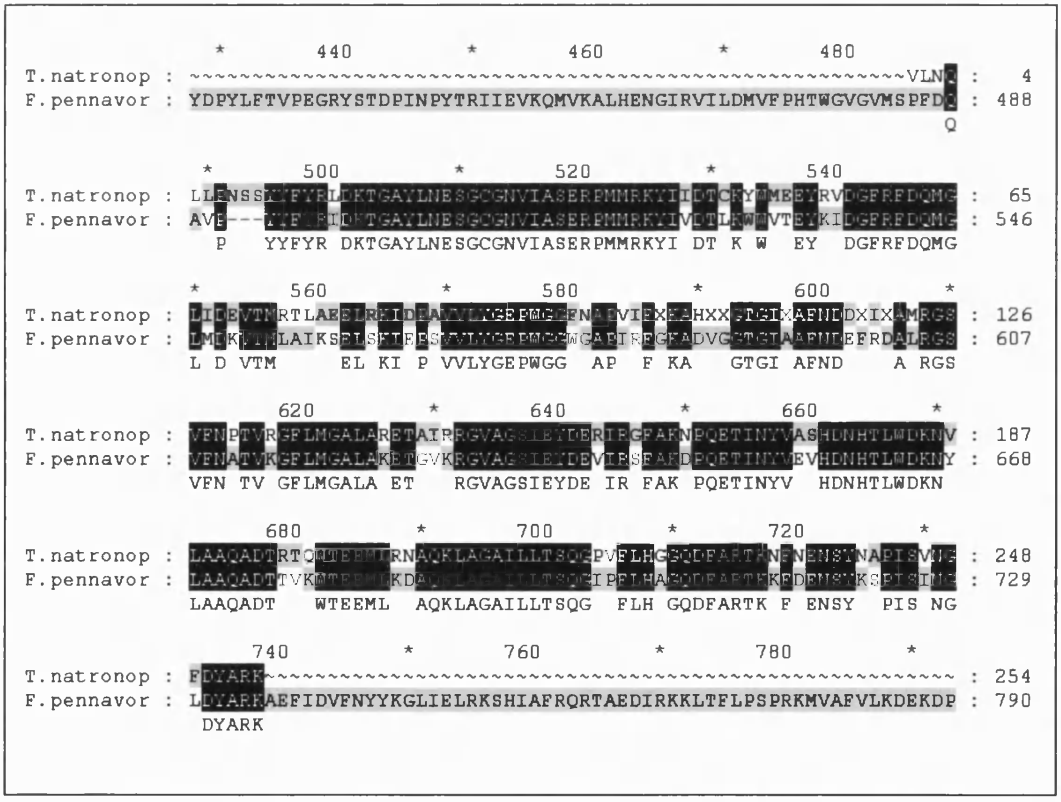


Fig. 4.10: The protein sequence alignment of PCR product amplified in Fig. 4.9 with pullulanase protein sequence of *F.pennavorans Ven*5.

Section 4: Identifying the Start and Stop Codons

Even though the majority of the *Thermopallium* pullulanase type I gene had been successfully obtained, the sequences containing the start and stop codons of the gene still remained elusive. Thus, several different attempts were launched to locate the start and the end of the gene. Amongst them were direct sequencing, inverse PCR, and PCR with biotinylated primers and paramagnetic beads. The first two approaches did not yield any results but the last method managed to identify the start and stop codons of *T.natronophilum* pullulanase gene.

4.8 Methods

4.8.1 Direct Sequencing with Specific Primers Using gDNA as Template

Reactions for DNA sequencing were carried out according to section 2.2.13. Reactions were set up with 3-5μg of gDNA and 10pmol of sequencing primer. The samples were made up to a final volume of 12μl with MQH_2O. Sequencing primers used are shown in Fig. 4.11.

Towards 5' End:

Direct Sequencing Reverse Primer

```
          K   D   D   F   G   V   V   A
(N) 5'-AAG GAC GAT TTT GGT GTT GTT GC-3'
(C) 3'-TTC CTG CTA AAA CCA CAA CAA CG-5'
(R) 5'-GC AAC AAC ACC AAA ATC GTC CTT-3'
     Degeneracy=1          Tm= 61.8°C
```

Towards 3'End:

Direct Sequencing Forward Primer

```
     N   A   P   I   S   V   N   G
5'-AAC GCA CCT ATT TCT GTT AAC GG-3'
     Degeneracy=1          Tm=59.5°C
```

Fig. 4.11: Specific primers used for direct sequencing of gDNA.

4.8.2 Identifying Start and Stop Codons with Inverse PCR

5U of *BamH*I, *Kpn*I, *Hind*III, *Eco*RI and *Sau*3AI were individually used to digest 1µg of gDNA in each enzymatic reaction. Reactions were set up as per the manufacturer's instructions. The final volume of each enzymatic reaction was 50µl. Digested DNA was then extracted with phenol/chloroform/isoamyl alcohol (24:1) using the method described in section 2.2.2. Digested DNA was then resuspended in 80µl of MQH_2O whereby 5µl of this was then used in the ligation reaction with 400U of T4 DNA Ligase (NEB, UK). Ligation reactions were set up as described in section 2.2.11 in a final volume of 20µl. 5µl of the ligation mixture was then used as template for inverse PCR. Specific primers used in the PCR reactions are given in Fig. 4.11 and Fig. 4.12. PCR was set up as described in section 2.2.6. The elongation step for all reactions was carried out at 72°C for 3 min, 5 min and 10 min. The combinations of primers used are as follows:

- Direct Sequencing Forward Primer/ Direct Sequencing Reverse Primer

  Annealing temperature = 60°C

- Direct Sequencing Reverse Primer/ FragA Forward

  Annealing temperature = 61°C

- Direct Sequencing Reverse Primer/ FragB Forward

  Annealing temperature = 60°C

- Direct Sequencing Forward Primer/ FragB Reverse

  Annealing temperature = 58°C

- Direct Sequencing Forward Primer/ FragC Reverse

  Annealing temperature = 58°C

FragA Forward

```
   D  G   K   A   E   V   W   L
5'-T GGA AAA GCG GAG GTG TGG C-3'
   Degeneracy: 1        Tm=64.1°C
```

FragB Forward

```
   Y   G   V   G   P   Q   S   A
5'- TAT GGT GTA GGT CCT CAA TCA GC-3'
   Degeneracy:1         Tm=61.7°C
```

FragB Reverse

```
     P   E   G   R   Y   A   T   D   P
(N) 5'-T CCA GAA GGT AGA TAC GCA ACA GAT C-3'
(C) 3'-A GGT CTT CCA TCT ATG CGT TGT CTA G-5'
(R) 5'-G ATC TGT TGC GTA TCT ACC TTC TGG A-3'
      Degeneracy=1         Tm=62.1°C
```

FragC Reverse

```
     L   R   K   I   D   P   A   V
(N) 5'-CTC AGA AAG ATT GAC CCT GCG GTA-3'
(C) 3'-GAG TCT TTC TAA CTG GGA CGC CAT-5'
(R) 5'-TAC CGC AGG GTC AAT CTT TCT GAG-3'
      Degeneracy=1         Tm=63.5°C
```

Fig. 4.12: Primers used for inverse PCR. Two other primers used in inverse PCR are found in Fig. 4.11.

4.8.3 Identification of Unknown DNA Flanking Sequences in 5' Terminal and 3' Terminal Regions Using Dynal® M-280 Streptavidin Paramagnetic Beads

The principles behind this technique will be discussed in the results section 4.9.3 in page 78.

4.8.3.1 First Round PCR with Flanking Primers and Pullulanase Gene Specific Primers

The first round of PCR was carried out using MasterTaq Kit (Eppendorf, Netherler-Hinz GmbH, Hamburg, Germany) with Specific Primer 1(SP1) and Flanking Primer 1 (FP1). The MasterTaq Kit is supplied with Taq DNA polymerase (5U/µl), 10× Taq Reaction Buffer with Mg^{2+} (500mM KCl, 100mM Tris-HCl, pH8.3 at 25°C, 15mM $MgCl_2$), and 5× TaqMaster solution. According to the manufacturer's manual, 5× TaqMaster improves the stability and efficiency of the Taq polymerase. It also makes the Taq polymerase less sensitive to PCR contaminants. This solution is found not to affect the annealing temperature or the melting behaviour of the DNA. PCR was carried out under 1 cycle of 96°C for 4 min, 30 cycles of 96°C for 1 min 15 sec, 60°C for 1 min 15s, 72°C for 3 min and finally 1 cycle of 72°C for 10 min. The PCR reaction was set up as follows:

Table 4.2: This shows the materials used in 1st round PCR of a 2step PCR reaction.

Component	Volume	Final Concentration
5× TaqMaster	**10µl**	**1×**
10× Taq Reaction Buffer	**5µl**	**1×(1.5mM $MgCl_2$)**
20mM dNTP-Mix	**1µl**	**0.4mM**
FP1 (10pmol/µl)	**1µl**	**0.2pmol/µl**
SP1 (10pmol/µl)	**1µl**	**0.2pmol/µl**
gDNA (50ng/µl)	**1µl**	**1ng/µl**
Taq DNA Polymerase (5U/µl)	**1µl**	**0.1U/µl**
MQH_2O	**30µl**	**-**
Total Volume	**50µl**	

- For 5'end:SP1 primer (10pmol/μl) / Flanking Primer (10pmol/μl).
- For 3'end:SPA primer (10pmol/μl) / Flanking Primer (10pmol/μl).

All primers used in this PCR can be found in Fig. 4.13 and their location on the pullulanase protein sequence can be found in Fig. 4.14.

4.8.3.2 Preparations of Dynal® M-280 Streptavidin Paramagnetic Beads

Dynabeads (Dynal®, Oslo, Norway) were first resuspended gently by inverting the vial upside down several times untill a homogenous solution was observed. 20μl of the suspension, containing 200μg of beads, was then transferred to a 1.5ml tube. The tube was then loaded onto a magnetic stand (Dynal MPC-E, Dynal®, Oslo, Norway) for 2 min to sediment the paramagnetic beads. The resulting supernatant was removed and the tube was then removed from the magnetic stand. Following this, the beads were washed with 40μl of 2× Binding and Washing buffer (10mM Tris-HCl, pH 7.5, 1mM EDTA, 2.0M NaCl, filtered with 0.22μm Milipore filter). The washing of beads with 2× Binding and Washing buffer was carried out twice and finally the beads were resuspended in 40μl of the same buffer to give a final concentration of 5μg/μl.

4.8.3.3 Immobilisation of DNA onto Dynal® M-280 Streptavidin Paramagnetic Beads

40μl of PCR reaction was added to 40μl of Dynabeads M-280 that were resuspended in 2× Binding and Washing buffer. This resulted in the beads being resuspended in a final concentration of 1× Binding and Washing buffer. The 1.5ml tube was then left rotating in a 1.5ml tube rotator overnight at room temperature. Overnight incubation is needed to facilitate binding of DNA of more than 1kb in size to the beads. Following the overnight incubation, the tube was removed and loaded onto the magnetic stand for 2 min. The resulting supernatant was removed. The beads were then washed thee times in 50μl of 1× Binding and Washing buffer (5mM Tris-HCl, pH 7.5, 0.5mM EDTA, 1.0M NaCl, filtered with 0.22μm Milipore filter). The supernatant was removed and the DNA on the beads was then melted.

4.8.3.4 Denaturing and Neutralising of Immobilised DNA duplex

Denaturation of DNA duplex was carried out by adding 8µl of 0.1M NaOH (filtered with 0.22µm Millipore filter) to the beads. Beads were left standing for 10 min, and the supernatant was then removed and neutralised with 4µl of 0.2M HCl and 1µl of 1M Tris-HCl, pH 8.0. 17µl of MQH_2O was added to the neutralised supernatant and 2µl was then used as template for the second PCR reaction.

4.8.3.5 Second Round PCR with Flanking Primers and Pullulanase Gene Specific Primers

Second round PCR was carried out as described in section 4.7.1. PCR conditions remains unchanged. Primers used were:

- For 5'end: TAG primer (10pmol/µl) / SP2 (10pmol/µl)
- For 3'end: TAG primer (10pmol/µl) / SPB (10pmol/µl)

All primers used in this PCR can be found in Fig. 4.13 and their location on the pullulanase protein sequence can be found in Fig. 4.14.

4.8.3.6 Breaking of the Streptavidin and Biotin Bond

To avoid solid-phase strand sequencing, the interaction between streptavidin and biotin has to be broken. This was carried out by adding 30µl of breaking buffer (3mM Tris-HCl, pH 8.0, 95% (v/v) formamide) and incubating the tube at 65°C for 30 min. The tube was then loaded onto the magnetic stand and the supernatant was removed as quickly as possible. Ethanol precipitation (section 2.2.12) was then employed to precipitate the DNA found in the supernatant and DNA was resuspended in 10µl of MQH_2O. Following this, 5µl of it was sequenced with 10pmol of primer SP3 (for 5'end) or SPC (for 3'end), which are annotated in Fig.4.14.

Flanking Primers:

FP1:	5′-CAG TTC AAG CTT GTC CAG GAA TTC $(N)_7$ GGA-3′
FP2:	5′-CAG TTC AAG CTT GTC CAG GAA TTC $(N)_7$ CCT-3′
FP3:	5′-CAG TTC AAG CTT GTC CAG GAA TTC $(N)_7$ GCA-3′
FP4	5′-CAG TTC AAG CTT GTC CAG GAA TTC $(N)_7$ GCT-3′
FP5:	5′-CAG TTC AAG CTT GTC CAG GAA TTC $(N)_7$ GGT-3′
FP6:	5′-CAG TTC AAG CTT GTC CAG GAA TTC $(N)_7$ CCA-3′
FP7:	5′-CAG TTC AAG CTT GTC CAG GAA TTC $(N)_7$ CGA-3′
FP8:	5′-CAG TTC AAG CTT GTC CAG GAA TTC $(N)_7$ CGT-3′
FP9:	5′-CAG TTC AAG CTT GTC CAG GAA TTC $(N)_7$ ATG-3′
FP10:	5′-CAG TTC AAG CTT GTC CAG GAA TTC $(N)_7$ GAA-3′
FP11:	5′-CAG TTC AAG CTT GTC CAG GAA TTC $(N)_7$ GAT-3′
FP12:	5′-CAG TTC AAG CTT GTC CAG GAA TTC $(N)_7$ AGA-3′
TAG:	5′-CAG TTC AAG CTT GTC CAG GAA TTC-3′

Specific Primers Towards 5'end of the Pullulanase Gene:

SP1:	5′-TTT GTT CAA CTC CTT GCA AGA GCC-3′
SP2:	5′-AAT GTC TAT GAA CCT GTC CAT TGC G-3′
SP3:	5′-CCA TTC ACC AAG TCT AAC AAT AAT ACC-3′

Specific Primers Towards 5'end of the Pullulanase Gene:

SPA:	5′-TGG GAC AAA AAC GTA TTA GCA GCT-3′
SPB:	5′-TCT CAA GGA CCA GTC TTT TTG CAC-3′
SPC:	5′-AAA ACT CGT ACA ACG CACCTA TTT C-3′
SPD:	5′-AGG TCC TTT ACG AAG TGA GCG G-3′
SPE:	5′-GAA ATT CTT GTG ATC TAC AAT GGG-3′

Fig. 4.13: Primers used in PCR reactions used to locate start and stop codons.

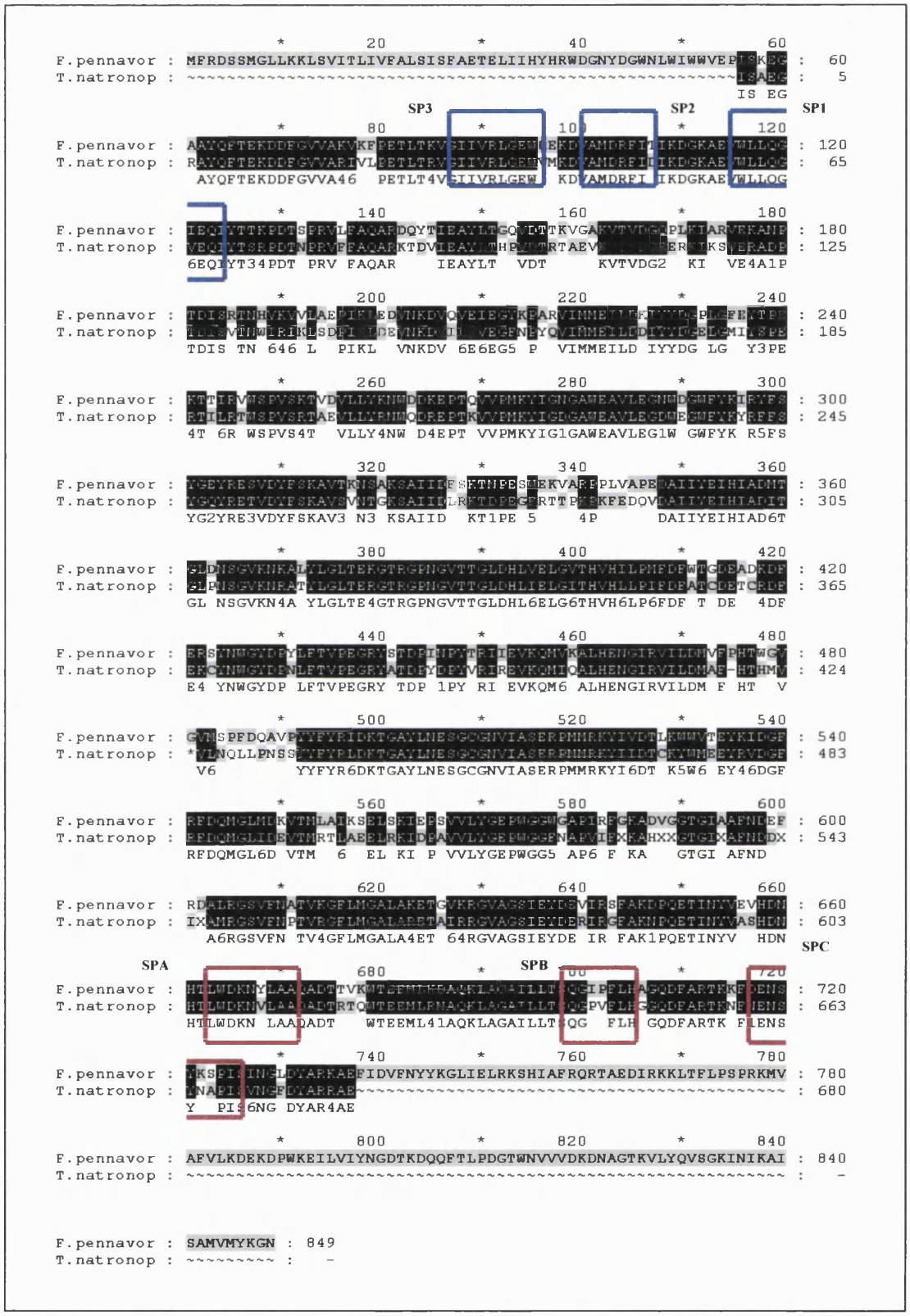


Fig.4.14: Incomplete pullulanase protein sequence. Peptide sequences boxed in blue were used to design primers for location of the start codon and peptide sequences boxed in red were used to design primers for location of the stop codon of the gene.

4.9 Results

4.9.1 Direct Sequecing from gDNA

Firstly, gDNA was directly sequenced using specific primers designed from the sequence data obtained from previous experiments. Towards the 5' end, a specific reverse primer was designed and towards the 3'end, a specific forward primer was designed. The locations of both primers on the *Thermopallium* pullulanase gene sequence are annotated in Fig. 4.16. Unfortunately, the primers failed to bind to the gDNA to give any sequences.

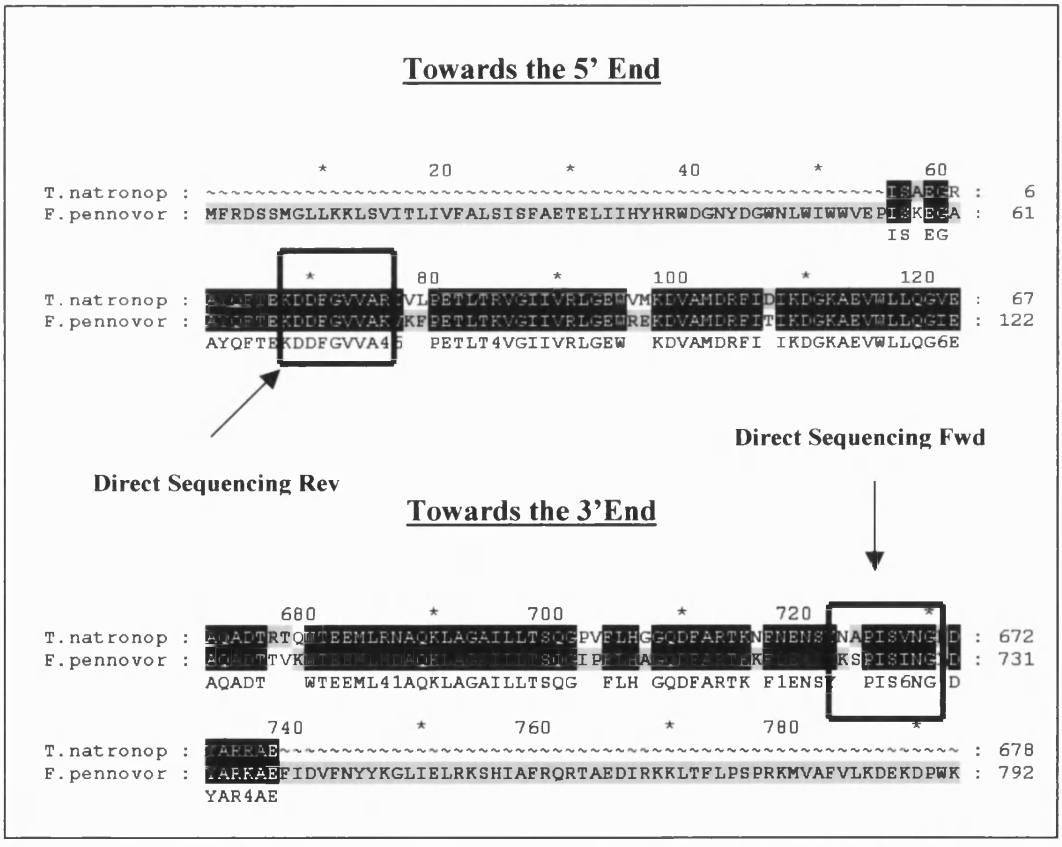


Fig. 4.15: Location of the specific primers used for direct sequencing of *Thermopallium* gDNA. Both primers were designed to read into the unknown region.

4.9.2 Inverse PCR

The second approach attempted was inverse PCR, whereby an unknown region next to known sequences can be amplified. This method involves the digestion of gDNA with different restriction enzymes and then circularising the gDNA. PCR was carried out using specific primers designed to read in opposite directions, which can then read into the unknown gene sequence. The positions of all primers used in inverse PCR relative to each other are annotated in Fig. 4.17.

HindIII cuts within the pullulanase gene sequence, with one restriction site approximately 860bp up stream of FragB Reverse primer (annotated as primer 3 in Fig. 4.16) and the second restriction site approximately 820bp down stream of FragB Forward primer (annotated as primer 4 in Fig. 4.16). No PCR products were successfully amplified using this approach.

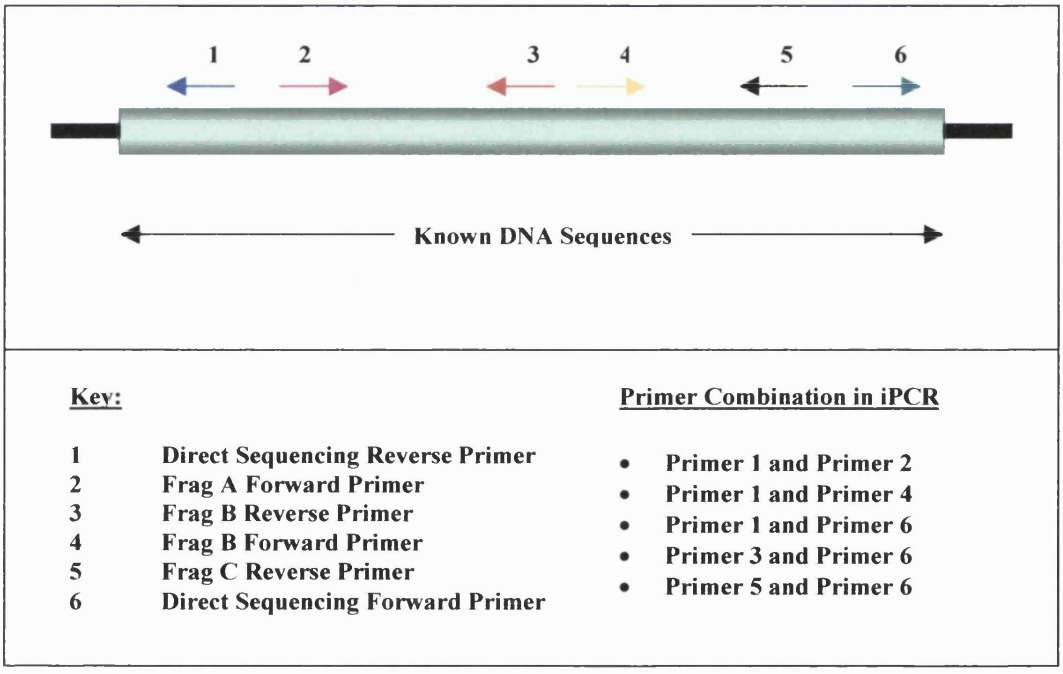


Fig. 4.16: The relative positions of primers used in inverse PCR.

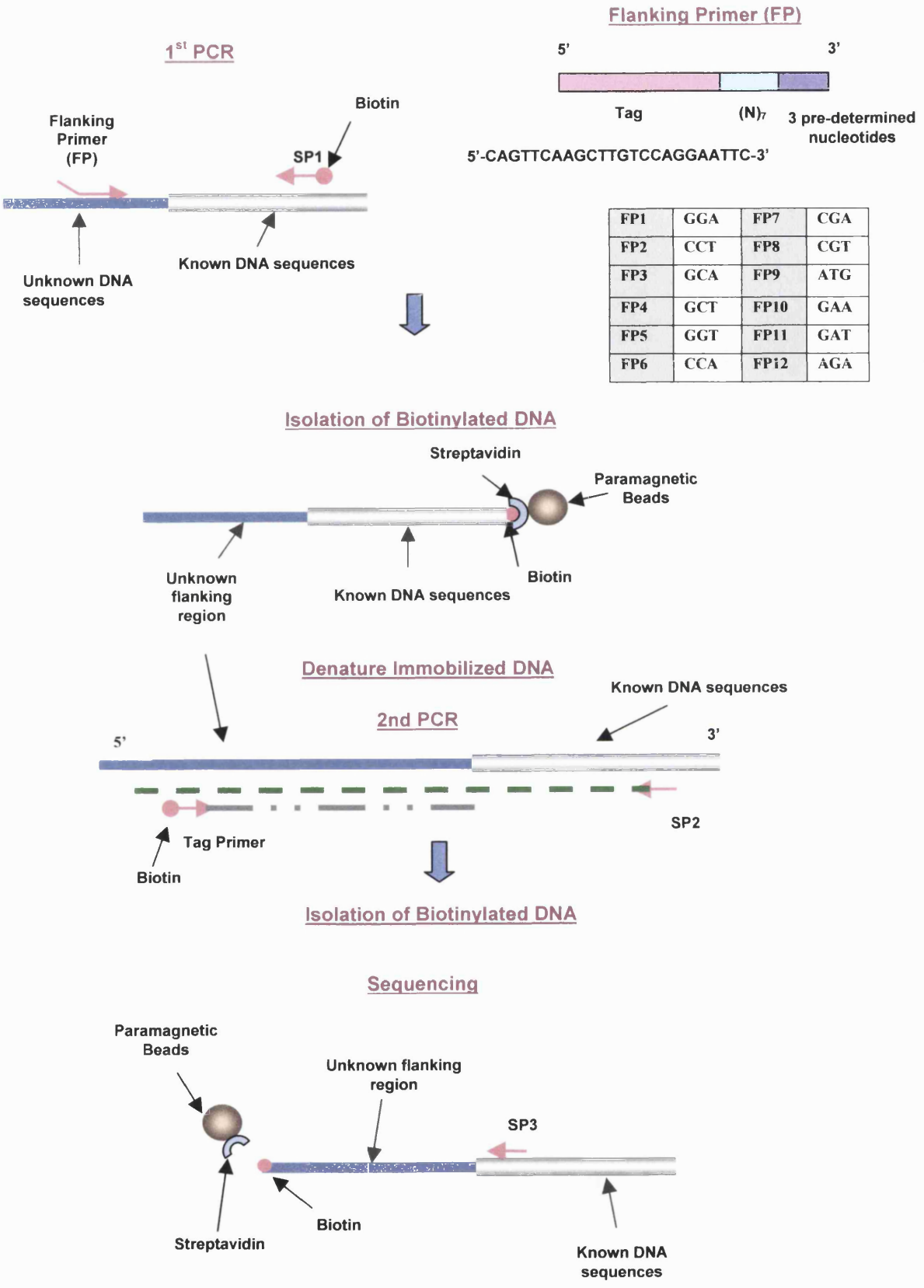


Fig. 4.17: Identification of the unknown flanking DNA regions using PCR with biotinylated primers and paramagnetic beads. Some primers used contain random hexamers.

4.9.3 Two Step PCR with Streptavidin Coated Paramagnetic Beads

In the search for the start codon, specific primers termed SP1, SP2 and SP3 were designed. All thee primers are reverse primers with SP1 positioned at the outermost, with SP2 and SP3 positioned upstream of it. The locations of the three primers are annotated in Fig. 4.14.

The first round of PCR (Fig. 4.17) involves primers SP1 and FP1. SP1 primer is biotinylated and is designed from the partial *Thermopallium* pullulanase gene sequence. Flanking primer 1(FP1) has a tag sequence towards the 5'end of the primer. This tag sequence contains 24 nucleotides that do not bind to the gDNA and will later be used to reamplify positive PCR products. The flanking primer also contains 7 random nucleotides and these nucleotides allow the flanking primer to anneal to any nucleotide sequences in the genome. In order to reduce the random annealing, following the hexamer are 3 predetermined nucleotides i.e. GGA for FP1. Incidentally, FP1 will only anneal to CCT sequences in the *Thermopallium* genome and the hexamer nucleotides in the FP1 primer will not discriminate any sequences located before the CCT nucleotides.

After the first round of PCR, products of various lengths will be obtained. These products are then immobilised on Dynal® M280 streptavidin paramagnetic beads via the biotin tag on the primer SP1. Only DNA fragments that have SP1 primer (biotinylated) incorporated will be immobilised on the beads. Following the immobilisation, the DNA duplex was melted. Attached to the paramagnetic beads will be the minus strand with the SP1 primer. The supernatant will contain the positive strand of the DNA with the Flanking Primers and it is this that will proceed to the next stage of PCR.

The second round of PCR involves the use of SP2 and TAG primers. TAG primer is of 24 nucleotides and is exactly the same sequence as the Tag sequence found in the Flanking Primers. Taq polymerase will transcribe all the way though from SP2 and will produce sequence complementary to the 24 nucleotides found in Flanking

Primers. The biotinylated TAG primers will now complement the sequence amplified from SP2 primers. Through this nested PCR, more of the desired PCR products are selected and amplified. Following this, the second round PCR products were immobilised on Dynal® M280 streptavidin paramagnetic beads.

Initially, an attempt to sequence the DNA attached to the beads were carried out but to no avail. Thus, to overcome this, the streptavidin and biotin bond were broken at 65°C for 30 min in 3mM Tris-HCl buffer, pH 8.0 containing 95% formamide (v/v). In total, 12 different Flanking Primers were designed (Fig.4.13). Some Flanking Primers yielded results and some did not. For the identification of the start codon for the pullulanase gene, FP10, FP11 primers were successful in identifying the ATG start site. It is important to have different flanking primers amplifying the same sequences albeit of different lengths. This shows the sequences obtained are true and original of the pullulanase sequence of *Thermopallium natronophilum*.

A possible Shine-Dalgarno sequence was also identified at –10 and –16 upstream of the ATG start site. A possible TATA box was also identified at position –46 upstream of the translation start site (Fig.4.18).

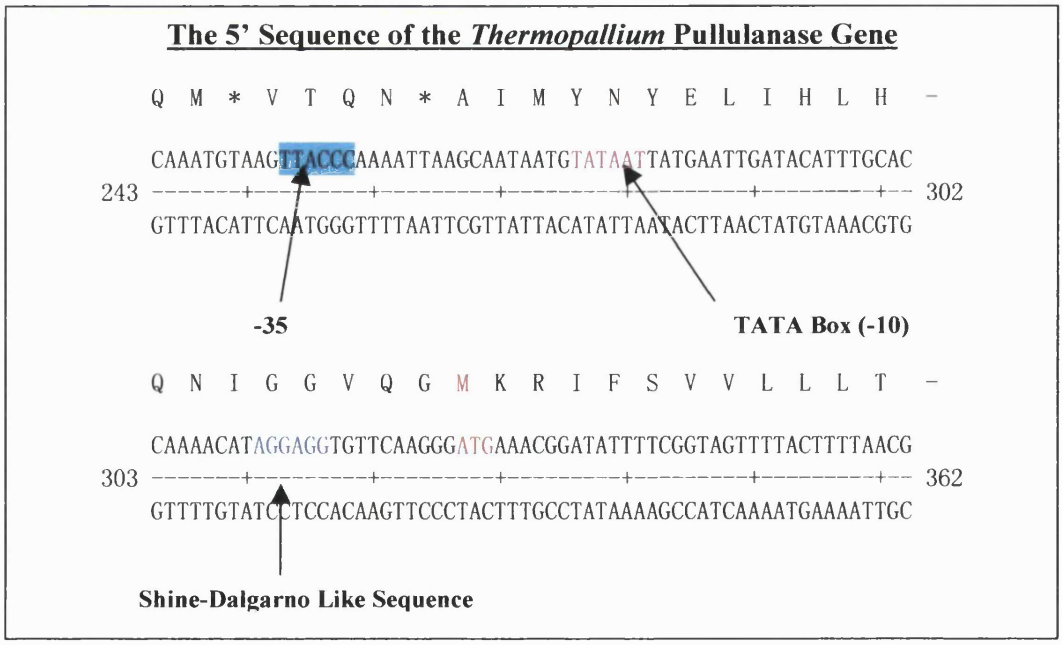


Fig. 4.18: The 5' sequence of the *Thermopallium* pullulanase gene. The start codon (M) is written in red and upstream of the ATG start site, a Shine-Dalgarno like sequence has been identified (blue). A potential TATA box (purple). and the –35 sequence (turquoise) had also been identified.4 The translated amino acid sequence is given above the DNA sequence.

A similar approach was carried out to identify the nucleotide sequence of the 3'end of the pullulanase gene. Three specific primers known as SPA, SPB and SPC were designed and their locations are annotated in Fig. 4.14. Initial PCR with FP1 and FP2 yielded more nucleotide sequence but still the stop codon was not obtained. To sequence right through the stop codon, two more specific primers known as SPD and SPE were designed from additional new sequences obtained from the initial trial and their positions are annotated in Fig. 4.19. The same experiments were repeated with FP1 and FP2 primers but SPD and SPE were used as the sequencing primers instead. In both cases, the 2 different primers managed to sequence right through the stop codon and sequences of approximately 90bp after the STOP codon was also obtained. This is shown Fig. 4.20.

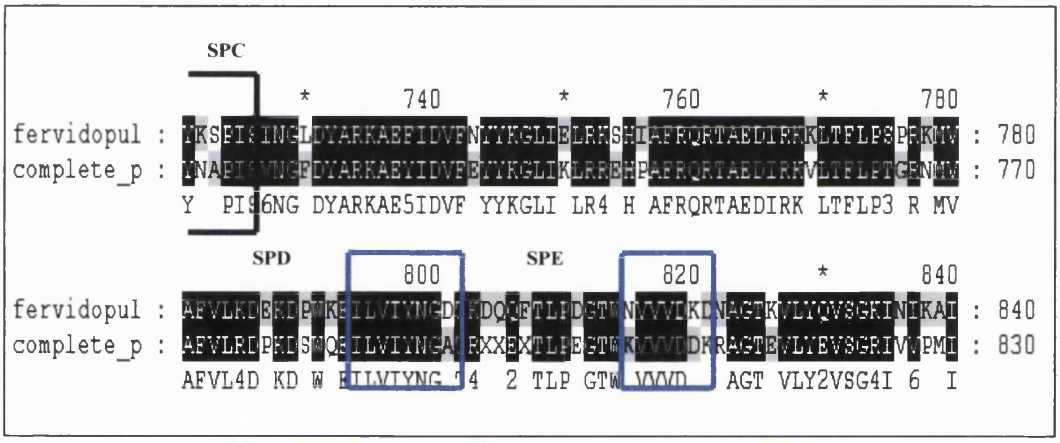


Fig. 4.19: Location of sequencing primers SPD and SPE.

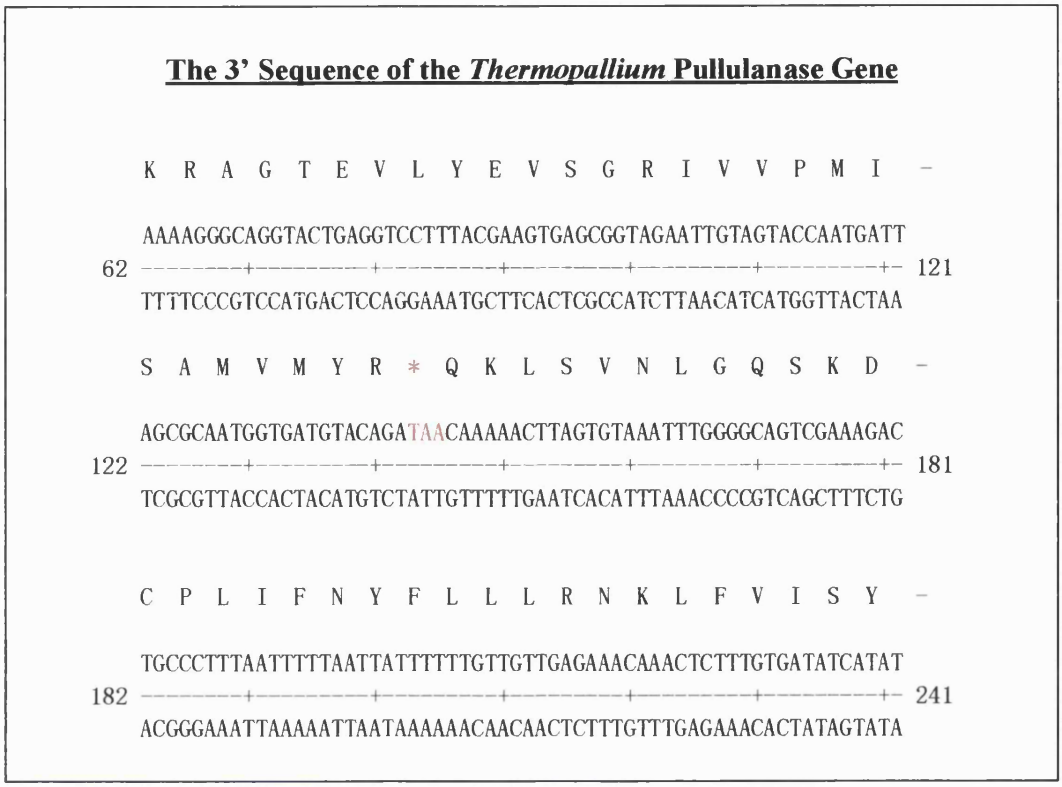


Fig. 4.20: The 3' sequence of the pullulanase type I gene identified by 2 step PCR with paramagnetic beads. The stop codon (*) is highlighted in red. The amino acid sequence is given above the DNA sequence.

In conclusion, the majority of the pullulanase gene sequence was obtained by PCR amplification with both specific and degenerate primers and the identification of start and stop codon using the two-step PCR with paramagnetic beads has been successful. The TATA box and the Shine-Dalgarno like sequence have also been identified. The full protein sequence is recorded in Fig. 4.21. The nucleotide and peptide sequence of the full length type I pullulanase can be found in Appendix I and the sequence alignment with type I pullulanase from *F.pennavorans* can be found in Appendix II.

Complete Protein Sequence of Pullulanase of *T. natronophilum*

```
  1  MKRIFSVVLL LTVVFLFAQT ELIIHYHRFD GNYEGWNLWI WYVEPISAEG
 51  RAYQFTEKDD FGVVARIVLP ETLTRVGIIV RLGEWVMKDV AMDRFIDIKD
101  GKAEVWLLQG VEQIYTSRPD TNPRVFFAQA RKTDVIEAYL THPVDTRTAE
151  VKVTVDGEER KIKSVERADP TDISVTNWIR IKLSDPIKLD EVNKDVILEV
201  EGFNPYQVIM MEILDDIYYD GELGMIYSPE RTILRTWSPV SRTAEVLLYR
251  NWQDREPTKV VPMKYIGDGA WEAVLEGDWE GWFYKYRFFS YGQYRETVDY
301  FSKAVSVNTG KSAIIDLRKT DPEGFRTTPK PKFEDQVDAI IYEIHIADIT
351  GLPNSGVKNR ATYLGLTERG TRGPNGVTTG LDHLIELGIT HVHLLPIFDF
401  ATCDETCRDF EKCYNWGYDP NLFTVPEGRY ATDPYDPYVR IREVKQMIQA
451  LHENGIRVIL DMVFPHTYGV GPQSAFDQTV PYYFYRLDKT GAYLNESGCG
501  NVIASERPMM RKYIIDTCKY WMEEYRVDGF RFDQMGLIDE VTMRTLAEEL
551  RKIDPAVVLY GEPWGGFNAP VRFGKAHVGG TGIGAFNDDF RDAMRGSVFN
601  PTVRGFLMGA LARETAIRRG VAGSIEYDER IRGFAKNPQE TINYVASHDN
651  HTLWDKNVLA AQADTRTQWT EEMLRNAQKL AGAILLTSQG PVFLHGGQDF
701  ARTKNFNENS YNAPISVNGF DYARKAEYID VFEYYKGLIK LRREHPAFRQ
751  RTAEDIRKVL TFLPTGRNMV AFVLRDPKDS WQEILVIYNG ATREQEFTLP
801  EGTWKVVVDD KRAGTEVLYE VSGRIVVPMI SAMVMYR
```

Fig. 4.21: The complete protein sequence for pullulanase type I of *Thermopallium natronophilum*.

Chapter 5: Analysis of the Pullulanase Type I Gene Sequence

5.1 Introduction

The aim of this chapter is to derive as much information as possible from the protein sequence obtained. Firstly, the protein sequence obtained was proven to be a pullulanase type enzyme. This is carried out by BLAST searching the protein sequence against the protein database. Through this approach, ten homologous proteins were successfully identified. All ten proteins are pullulanase type enzymes, all belonging to glycosyl hydrolase family 13.

In addition, the three domains that define the members of the glycosyl family 13 were also identified on the protein sequence from *T.natronophilum* and the four homologous regions found in the glycosyl hydrolase family 13 were then highlighted.

An attempt to predict the TIM barrel secondary structure was carried out by sequence alignments with known or predicted protein structures using web-based programmes. Identification of a conserved region that specifically discriminates pullulanase type I from other pullulanases was also attempted.

Following this, an investigation of possible amino acid residues that are important in catalytic activity and substrate binding was also carried out. Last but not the least, a signal peptide for this enzyme was also identified using a web-based programme. This was important for the next step of the project, namely the cloning and expression of the recombinant protein.

Section 1: Identification of Homologous Proteins Using Blast Search

5.2 Methods

5.2.1 Identifying Homologous Proteins in the Sequence Database

Proteins homologous to the pullulanase sequence from *Thermopallium* were searched using BLASTp. Ten homologous proteins were then aligned with the protein sequence from *T.natronophilum* using GeneDoc Version 2.6.001 (Nicholas *et al.*, 1997)

5.3 Results

5.3.1 Identifying Homologous Protein Using BLASTp

All the proteins identified by BLASTp that share homology with *Thermopallium* type I pullulanase belong to the glycosyl hydrolase family 13. From protein sequence alignments, the ten proteins with the greatest homology to *Thermopallium* type I pullulanase were identified, and are listed in Table 5.1.

From the percentage identity of the protein sequences obtained, the nucleotide sequence amplified from the *Thermopallium* gDNA is indeed a gene encoding a type I pullulanase with the highest identity with type I pullulanase from *F.pennavorans Ven*5 (74%), followed by *T.maritima* (65%). This is in good agreement with the phylogenetic tree based on 16S rRNA gene sequences (Fig. 1.2), with *F.pennavorans Ven*5 being the closest relative to *T.natronophilum,* followed by *T.maritima.*

As mentioned in Chapter 1, members of the glycosyl hydrolase family 13 all contain four highly conserved regions. The four regions identified in the alignment are shown in Table 5.2 and annotated as RI, RII, RIII and RIV. An additional highly conserved region that is normally found in type I pullulanase has also been highlighted as RV in Table 5.2.

The strictly conserved sequences in pullulanase type I from *T.natronophilum*, *F.pennavorans Ven5* and *T.maritima* are DMVFPH (RI), DGFRFDQMGL (RII), EPWG (RIII) and YNWGYDP (RV). In RIV, only Y---NH are strictly conserved. The effects of the differences between the different enzymes have yet to be determined experimentally.

Table 5.1: Results obtained from Blastp search. These results show the 10 pullulanases sharing significant sequence identity with that from the *T.natronophilum* enzyme.

Organisms	Enzyme	Identity
Fervidobacterium pennivorans	Pullulanase type I	74%
Thermotoga maritima	Pullulanase type I	65%
Geobacillus stearothermophilus	Pullulanase	43%
Bacillus cereus ATCC 14579	Pullulanase	41%
Clostridium perfringens str. 13	Pullulanase	41%
Thermus sp. IM6501	Pullulanase	41%
Bacillus anthracis	Pullulanase	41%
Bacteroides thetaiotaomicron VPI-5482	Pullulanase	40%
Bacillus halodorans	Pullulanase	39%
Bacillus subtilis	Pullulanase	39%
Bacillus halodurans	Pullulanase	39%

Table 5.2: The four highly conserved regions (RI-RIV) that define the α-amylase family. RV is highly conserved in type I pullulanase. All the ten enzymes are identified as homologous proteins to *T.natronophilum* pullulanase via BlastP search. * indicates the amino acid that is conserved in all strains.

Organisms	RI (β3)	RII (β4)	RIII (β5)	RIV(β7)	RV
Thermopallium natronophilum	DMVFPH	DGFRFDQMGL	EPWG	YVASHD	YNWGYDP
Fervidobacterium pennivorans	DMVFPH	DGFRFDQMGL	EPWG	YVEVHD	YNWGYDP
Thermotoga maritima	DMVFPH	DGFRFDQMGL	EPWG	YAACHD	YNWGYDP
Geobacillus steareothermophilus	DVVYNH	NGFRFDLMGI	EGWD	YVESHD	YNWGYNP
Bacillus cereus ATCC 14579	DVVYNH	DGFRFDLMGI	EGWD	YVEAHD	YNWGYDP
Clostridium perfringens str. 13	DVVYNH	DGFRFDLMAL	EGWN	YCSAHD	YNWGYDP
Thermus thermophilus HB8	DAVYNH	DGFRFDLMGV	EGWD	YVECHD	YNWGYNP
Bacillus anthtracis	DVVYNH	DGFRFDLMGI	EGWD	YVEAHD	YNWGYDP
Bacteroides thetaiotaomicron VPI-5482	DVVYNH	DGFRFDLMGI	EGWV	FVSCHD	YNWGYDP
Bacillus halodorans	DVVFNH	DGFRFDLMGI	EGWD	YVEAHD	YNWGYDV
Bacillus subtilis	DVVFNH	DGFRFDLLGI	EGWD	YVESHD	YNWGYNP
	* * *	*****	* *	**	*****

Section 2: Identification of Domains in *Thermopallium* Type I Pullulanase

5.4 Methods

Domain identification was carried out using Protein Families Database of Alignment and HMMs (Pfam) (www.sanger.ac.uk/Software/Pfam/) (Bateman *et al.*, 2002)

5.5 Results

Protein domains can be defined as basic units of protein structure consisting of structural folds with defined functions. Identification of protein domains may provide an insight into the function of a protein. Protein domains are established based on available protein structures and their protein sequences. Protein domains can be computationally identified by analysing databases of these known structures and observing the conserved region posed by a certain pattern of sequences.

Three domains of type I pullulanase of *Thermopallium* have been identified with Pfam, the first domain being the bacterial pullulanase-associated domain (PUD) spanning residues 19-123. This domain is mainly found in pullulanase or carbohydrate debranching enzyme, and can be found either on the C terminal or the N terminal side of the α-amylase catalytic domain. Other enzymes that were aligned to this domain include type I pullulanase from *F.pennavorans Ven5* and *T.maritima*; pullulanase precursor from *K.pneumoniae*; amylopullulanase and pullulanase from *S.pneumoniae*; alkaline amylopullulanase from *B.halodurans*, *Bacillus sp.* KSM-1876, *B.cereus* and *B.anthracis.*

The second domain identified is the isoamylase N-terminal N domain, spanning residues 222-302. This domain is commonly found in debranching enzymes. Some of the enzymes with this domain are pullulanase from *Geobacillus thermoleovorans*; amylopullulanase from *B.halodurans*, *Bacillus sp.* KSM-1876, *B.cereus* and *B.anthraci*; type I pullulanase from *F.pennavorans Ven5* and *T.maritima.* In amylase

from *Thermoactinomyces vulgaris* R-47 and neopullulanase from *B.stearothermophilus*, the N terminal in each enzymes were involved in dimer formation. Isoamylase of *Pseudomonas amyloderamosa* is a monomer and some residues in its N domain is found to form an interface with domain A.

The last domain identified is the α-amylase domain, spanning residues 362-742. This domain bears the $(\alpha/\beta)_8$ structure, a definite characteristic of all members of the glycosyl hydrolase family 13. Other enzymes carrying this domain are cyclomaltodextrinase from *T.volcanium*; α-amylase from *P.furiosus*; maltodextrin glycosyltransferase, α-amylase, pullulanase from *T.maritima* and *Bacillus sp.*

Section 3: Identification of the α-helices and β-sheets in Domain A, B and C of the Pullulanase Protein Sequence

5.6 Methods

5.6.1 α-helices and β-sheets Prediction in Domain A

Several different analyses were carried out for prediction of the $(\alpha/\beta)_8$ structure. These involve the alignment of the pullulanase protein sequence of *T.natronophilum* with the neopullulanase protein sequence of *Bacillus stearothermophilus*, (Hondoh *et al.*, 2003), the latter enzyme having a known crystallographic structure (PDB ID: 1j0j).

The second strategy used was to align the *Thermopallium* pullulanase protein sequence against the pullulanase protein sequences of *K. aerogenes* W70, *B.stearothermophilus* TRS128, *Thermus Sp.* AMD-33, *B.acidopullulyticus*, *C.saccharolyticus* and *T.maritima*. The $(\alpha/\beta)_8$ structures of all the enzymes listed were previously predicted by Kashiwabara *et al.* (1999). Structural predictions were carried out using solved secondary structures of α-amylase from *A.oryzae*; isoamylase from *P.amyloderamosa* JD210 and oligo-1,6-glucosidase from *B.cereus* ATCC 7064 as templates.

The third strategy carried out was based on a web-based method known as Biology WorkBench Version 3.2 (http://workbench.sdsc.edu/). This uses a programme called PELE, which groups together the predictions carried out by seven different programmes (namely, **BPS**: Burgess *et al.* (1974); **D_R**: Dele`age. and Roux (1987); **DSC**: King and Sternberg (1996); **GGR**: Garnier *et al.* (1996); **GOR**: Garnier *et al.* (1978); **H_K**: Holley and Karplus (1989); **K_S**: King and Sternberg (1990). Data from these seven programmes are then grouped together under a function known as ***JOI*** (Joint prediction), whereby secondary structure prediction is made by drawing a conclusion based on the data generated. Thus for example, for a helix to be predicted,

almost all seven programmes must give the same result in order for this to be assigned.

Data collected from these three strategies were then superimposed on each other to obtain the most plausible prediction for *Thermopallium* pullulanase.

5.6.2 Prediction of Domain B: Loop Connecting β_3 to α_3

For the prediction of Domain B, the protein sequence from *Thermopallium* pullulanase was aligned against pullulanase protein sequence of *K. aerogenes* W70, *B.stearothermophilus* TRS128, *Thermus Sp.* AMD-33, *B.acidopullulyticus*, *C.saccharolyticus* and *T.maritima*. The $(\alpha/\beta)_8$ structures of all the enzymes listed were previously predicted by Kashiwabara *et al.* (1999).

Computational prediction using PELE was also carried out. Data presented in neopullulanase protein sequence of *Bacillus stearothermophilus*, (Hondoh *et al.*, 2003) were also considered.

5.6.3 α-helices and β-sheets Prediction of Domain C

Prediction of Domain C was carried out by aligning the pullulanase sequence from *Thermopallium* with neopullulanase protein sequence of *Bacillus stearothermophilus*, (Hondoh, *et al.*, 2003). Computational prediction using PELE was also carried out.

5.7 Results

5.7.1 Identifying the α-helices and β-sheets in Domain A

In this section, the predicted eight α-helices and eight β-sheets based on aligning the *Thermopallium* protein sequence against neopullulanase of *B.stearothermophilus* (Hondoh *et al.*, 2003) are tabulated in Table 5.3.

The predicted peptide sequences contributing to the TIM-barrel of pullulanase protein of *T.natronophilum* obtained via alignment with predicted data carried out by Kashiwabara *et al.* (1999) are included in Table.5.4.

The data obtained from prediction using the computational approach, PELE were annotated Fig. 5.1. A residue that is predicted to be in a α-helix has been annotated H; β-helix an E, and coils C. Each of the programmes gave rise to different predictions, but predictions in β-sheets and some α-helices do coincide with each other.

For data analysis, when an α-helix residue is predicted adjacent to a β-sheet residue without being separated by at least a coil, the prediction will not be regarded as robust, due to the reason that each residue will have only a 50% chance of being correctly predicted.

In order to give the best overall structure prediction, all three results obtained from different strategies were aligned together in Fig. 5.1. Two identical protein sequences of type I pullulanase from *T.natronophilum* were annotated in this figure. The predicted structural assignments using the first method (i.e. comparison with neopullulanase from *B.stearothermophilus*) were charted in protein sequence (1). On the other hand, the predicted structural assignments obtained by alignment with protein sequences from type I pullulanase of *K.aerogenes* W70, *B.stearothermophilus* TRS 128, *Thermus Sp.* AMD-33, *B.acidopullulyticus*, *C.saccharolyticus* and *T.maritima* were charted on protein sequence (2).

Table 5.3: Predicted amino acid sequences contributing to the $(\alpha/\beta)_8$ of domain A in *T.natronophilum* pullulanase. This is done by aligning the protein sequence with protein sequence of neopullulanase from *B.stearothermophilus* (Hondoh *et al.*, 2003).

α/β	Neopullulanase of *B.stearothermophilus*	Pullulanase of *T.natronophilum*
β1	^{136}VWYQIF^{141}	^{340}IIYEIH^{345}
β2	^{192}GIYL^{195}	^{391}HVHL^{394}
β3	^{237}IRVMLDA^{243}	^{456}IRVILDM^{462}
β4	^{324}GWRL^{327}	^{529}GFRF^{532}
β5	^{354}ILGEI^{358}	^{559}LYGEP^{563}
β6	^{374}V (Incomplete)	^{584}M
β7	^{417}FNL^{419}	^{642}INY^{644}
β8	^{454}CI^{455}	^{693}FL^{694}
α1	^{174}DLQGIIDHLDYLVD^{187}	^{375}NGVTTGLDHLIE^{386}
α2	^{222}KETLKTLIDRCHE^{234}	^{441}IREVKQMIQALHE^{453}
α3	^{303}PEVKRYLLDVATYWIR^{318}	^{508}PMMRKYIIDTCKYWME^{523}
α4	^{335}HEFWREFRQEVKAL^{348}	^{540}EVTMRTLAEELRKI^{553}
α5	^{363}MP^{364}	^{569}AP^{570}
α6	^{377}YPFTDGVLR^{385}	^{588}DDFRDAMR^{594}
α7	^{394}ARQFANQMMHVLH^{406}	^{618}RRGVAGSIEYDERIR^{632}
α8	^{410}NNVN^{414}	^{636}KNPQ^{639}

Table 5.4: Predicted peptide sequences that give rise to the $(\alpha/\beta)_8$ TIM barrel structure (Domain A) in each pullulanase. This prediction was carried out by Kashiwabara *et al.* (1999), using the known secondary structure of α-amylase from *A.oryzae*, isoamylase from *P.amyloderamosa* JD210, and oligo-1,6-glucosidase from *B.cereus* ATCC 7064 as templates. Note: The structures of the enzymes from *A.oryzae*, *P.amyloderamosa* JD210, *B.cereus* ATCC 7064 were solved using x-ray crystallography.

Organism	β1	α1	β2	α2	β3
K.aerogenes W70	MTIHES	MVQHLKQLSAS	HIELL	IKE-FRTMIQAIKQD	MNVIMDVV
B.stearothermophilus TRS 128	AVIYEM	-PTGLSQLKEL	HVELL	IRE-LKRAIHT-LQS	IRVIMDVV
Thermus Sp. AMD-33	AIIYEL	TATGLSYVKEL	HVQLM	IVE-LKQAIHT-LHE	LRVVMDAV
B.acidopullulyticus	-VIYEV	VKTGIDSLKEL	AVQLQ	ITQ-LKQLIQS-IHK	IAINMDVV
C.saccharolyticus	AIIYEM	ISTGLLHLKEL	HVHLL	LKE-LRTMIKK-LHE	IGVVMDVV
T.maritima	AIIYEI	VTTGLSHLVEL	HVHIL	IRE-VKEMVK-ALHK	IGVIMDMV
T.natronophilum	AIIYEI	VTTGLDHLIEL	HVHLL	IRE-VKQMIQ-ALHE	IRVILDMV

Organism	α3	β4	α4	β5	α5
K.aerogenes W70	RMFAKLIADSLAVWTT	GFRFDL	-AQILSAW-ER	IYF-FGE	EIASQINL
B.stearothermophilus TRS 128	KMVRKWIIDSVRFWVE	GFRFDL	VEHM-KAVREM	ILV-FGE	-ATMQNAE
Thermus Sp. AMD-33	RMARRWIVDSVVFWAK	GFRFDL	IEHM-KAVRDA	ILV-YGE	--RWSNA-
B.acidopullulyticus	RMVQKFVLDSVKYWVK	GFRFDL	-DHMA-KISKE	ILV-YGE	LVTKGQQK
C.saccharolyticus	PMVRKFILDTIIYWTE	GFRFDL	VKQS-RQVANE	ALV-YGE	EMATIL--
T.maritima	PMMRKFIVDTVTYWVK	GFRFDQ	KKTML-EVERA	IIL-YGE	RFGKSDVA
T.natronophilum	PMMRKYIIDTCKYWME	GFRFDQ	EVTMRTL-AEE	VVL-YGE	RFGKAHVG

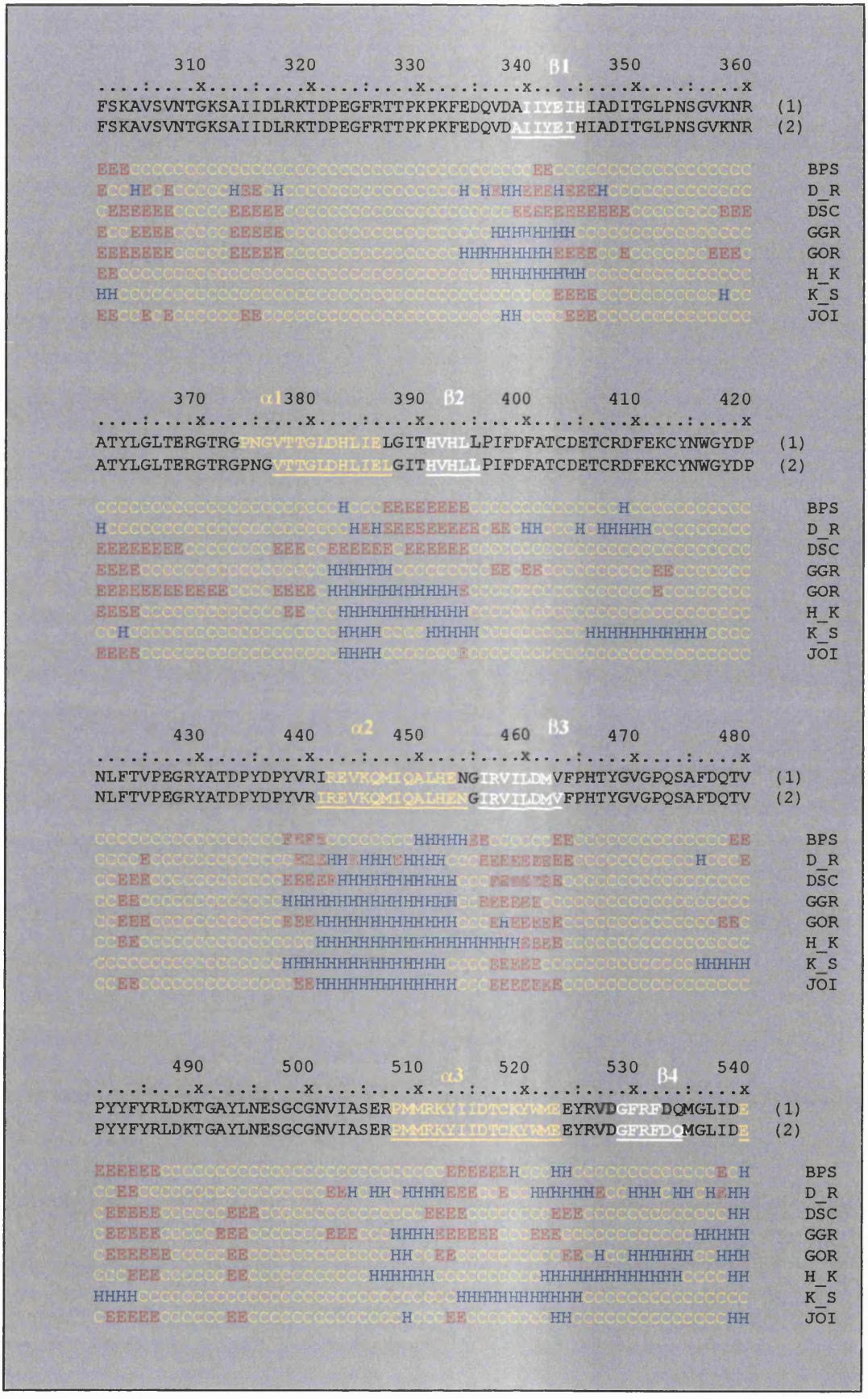
Organism	β6	α6'	α6	β7	α7	α7'
K.aerogenes W70	GTF	LRDAVR	ARHLAD-LTRLGMA	V-VNY	RVRMQAVSLATVM	LWDMIS
B.stearothermophilus TRS 128	GYF	FRDSAK	-EKVK-VAISGSI-	-TINY	RKKRQKLATAIVL	FWDKME
Thermus Sp. AMD-33	AYF	FRDAVK	-EQVK-LAIAGSL-	-SINY	RRKRQKLATAIVL	FWDKME
B.acidopullulyticus	GVF	IRNGLD	VNVIK-NRVMGSI-	-TINY	RIKMDELAQAVVF	LWDKIS
C.saccharolyticus	GLF	IRDAIR	VGRLK-QGIKAAI-	C-VNY	IDRATRLANAIVL	LFDKLQ
T.maritima	AAF	FRDAIR	YGKE--TKIKRGVV	-TINY	LKNAQKLAGAILL	LWDKNY
T.natronophilum	GAF	FRDAMR	AR-E--TAIRRGVA	-TINY	LRNAQKLAGAILL	LWDKNV

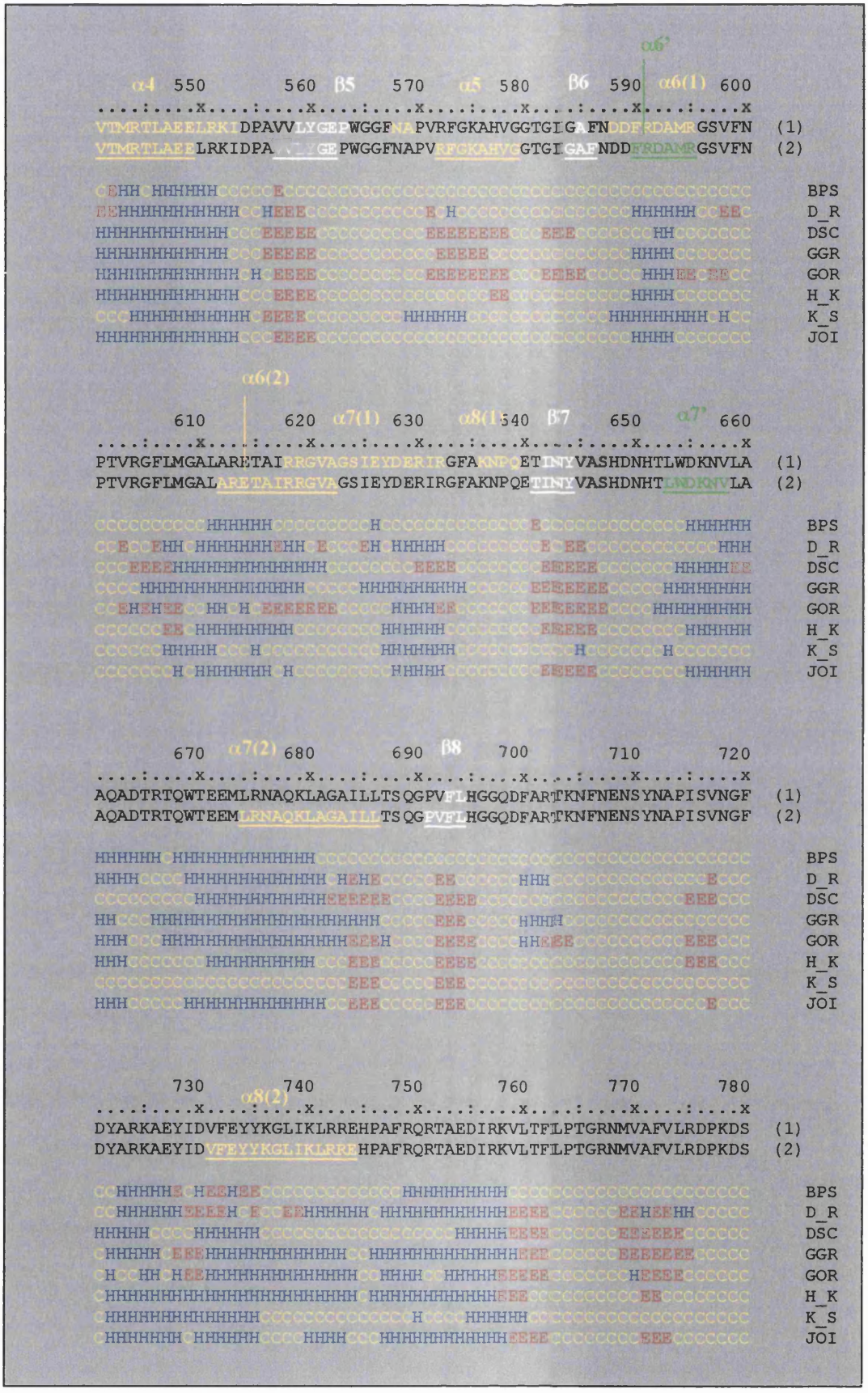
Organism	β8	α8
K.aerogenes W70	IAFDQ	-----YD-IIARVK---
B.stearothermophilus TRS 128	IPFL-	V-RYVQG-LI-RI-RKS
Thermus Sp. AMD-33	NPFL-	--VRLR---IDPL-RRA
B.acidopullulyticus	VPFMQ	VFDYYSW-LI-HL-RDN
C.saccharolyticus	VAFL-	-FKFYCD-LI-NL-RRK
T.maritima	VPFL-	VFNYHKG-LI-KL-RKE
T.natronophilum	PVFL-	VFEYYKG-LI-KL-RRE

Fig 5.1: This is the cumulative predictions for Domain A of *Thermopallium* pullulanase. The protein sequence is the protein sequence of *Thermopallium* pullulanase. Both sequences are identical with the numbered sequence being:

(1) Secondary Structures predicted from alignment with neopullulanase *B.stearothermophilus*.
(2) Secondary Structures predicted from alignment with type I pullulanase of *K.aerogenes* W70.

The rest of the predictions were carried out by the computer web-based programme known as PELE which groups together the predictions carried out by seven different programmes. These programmes BPS: Burgess *et al.* (1974); D_R: Dele`age. and Roux (1987); DSC: King and Sternberg (1996); GGR: Garnier *et al.* (1996); GOR: Garnier *et al.* (1978); H_K: Holley and Karplus (1989); K_S: King and Sternberg (1990). Data from these seven programmes were then grouped together under a function known as *JOI,* Joint prediction, whereby secondary structures were predicted. α-helix has been annotated H; β-helix an E, and coils C.


310 320 330 340 β1 350 360
FSKAVSVNTGKSAIIDLRKTDPEGFRTTPKPKFEDQVDAIIYEIHIADITGLPNSGVKNR (1)
FSKAVSVNTGKSAIIDLRKTDPEGFRTTPKPKFEDQVDAIIYEIHIADITGLPNSGVKNR (2)
BPS
D_R
DSC
GGR
GOR
H_K
K_S
JOI
370 α1 380 390 β2 400 410 420
ATYLGLTERGTRGPNGVTTGLDHLIELGITHVHLLPIFDFATCDETCRDFEKCYNWGYDP (1)
ATYLGLTERGTRGPNGVTTGLDHLIELGITHVHLLPIFDFATCDETCRDFEKCYNWGYDP (2)
BPS
D_R
DSC
GGR
GOR
H_K
K_S
JOI
430 440 α2 450 460 β3 470 480
NLFTVPEGRYATDPYDPYVRIREVKQMIQALHENGIRVILDMVFPHTYGVGPQSAFDQTV (1)
NLFTVPEGRYATDPYDPYVRIREVKQMIQALHENGIRVILDMVFPHTYGVGPQSAFDQTV (2)
BPS
D_R
DSC
GGR
GOR
H_K
K_S
JOI
490 500 510 α3 520 530 β4 540
PYYFYRLDKTGAYLNESGCGNVIASERPMMRKYIIDTCKYWMEEYRVDGFRFDQMGLID (1)
PYYFYRLDKTGAYLNESGCGNVIASERPMMRKYIIDTCKYWMEEYRVDGFRFDQMGLID (2)
BPS
D_R
DSC
GGR
GOR
H_K
K_S
JOI


α4 550 560 β5 570 α5 580 β6 590 α6(1) 600 α6'
VTMRTLAEELRKIDPAVVLYGEPWGGFNAPVRFGKAHVGGTGIGAFNDDFRDAMRGSVFN (1)
VTMRTLAEELRKIDPA_VLYGEPWGGFNAPVRFGKAHVGGTGIGAFNDDFRDAMRGSVFN (2)
BPS
D_R
DSC
GGR
GOR
H_K
K_S
JOI
610 620 α7(1) 630 α8(1) 640 β7 650 α7' 660 α6(2)
PTVRGFLMGALARETAIRRGVAGSIEYDERIRGFAKNPQETINYVASHDNHTLWDKNVLA (1)
PTVRGFLMGALARETAIRRGVAGSIEYDERIRGFAKNPQETINYVASHDNHTLWDKNVLA (2)
BPS
D_R
DSC
GGR
GOR
H_K
K_S
JOI
670 α7(2) 680 690 β8 700 710 720
AQADTRTQWTEEMLRNAQKLAGAILLTSQGPVFLHGGQDFARTKNFNENSYNAPISVNGF (1)
AQADTRTQWTEEMLRNAQKLAGAILLTSQGPVFLHGGQDFARTKNFNENSYNAPISVNGF (2)
BPS
D_R
DSC
GGR
GOR
H_K
K_S
JOI
730 740 α8(2) 750 760 770 780
DYARKAEYIDVFEYYKGLIKLRREHPAFRQRTAEDIRKVLTFLPTGRNMVAFVLRDPKDS (1)
DYARKAEYIDVFEYYKGLIKLRREHPAFRQRTAEDIRKVLTFLPTGRNMVAFVLRDPKDS (2)
BPS
D_R
DSC
GGR
GOR
H_K
K_S
JOI

5.7.2 Discussion (Domain A Prediction)

The diagrammatic topology of the domain A of *B.stearothermophilus* neopullulanase has been annotated in Fig. 5.2. In this domain, β1 is followed by α1. This alternating β/α pattern continues till β6 and α6. From α6 to β7, there are two α-helices, annotated as α7 and α8 in Fig. 5.2. After β7, two additional α-helices were reported and annotated α9 and α10 before coming to β8 i.e. the last β-sheet that forms the TIM barrel.

Neopullulanase is not a type I pullulanase and members of the α-amylase family are notorious for having less than 30% homology in sequences. Due to these reasons, the predicted amino acids contributing to the α-helices and β-sheets might vary to a certain degree from the true structure of the type I pullulanase. Despite this, these predicted sequences can still be used as a starting point and be used to compare with data generated by different approaches. All the residues discussed in the following paragraphs are in *T.natronophilum* numbering. A summary of the predicted sequences based on data generated by these three approaches is summarised in Table 5.5.

For the predictions of the β1 to β7 sheets and α1 to α4 helices, the predictions by the three approaches correspond well with each other but the predictions of α6, α7 and α8 are not as straight forward. The data generated by the three different approaches will be discussed in the following paragraphs.

α6 of *B.stearothermophilus* neopullulanase is homologous to an additional helix identified as α6' in domain A of α-amylase from *A.oryzae*, *P.amyloderamosa* JD210 and oligo-1,6-glucosidase from *B.cereus* ATCC 7064 (Table 5.4). In pullulanase of *T.natronophilum*, this stretch of amino acids corresponds to ^{588}DDFRDAMR^{594}. In the case for type I pullulanase from *T.natronophilum*, this stretch of amino acids might be α6 or α6'. PELE has also determined these eight amino acids to be α-helix residues.

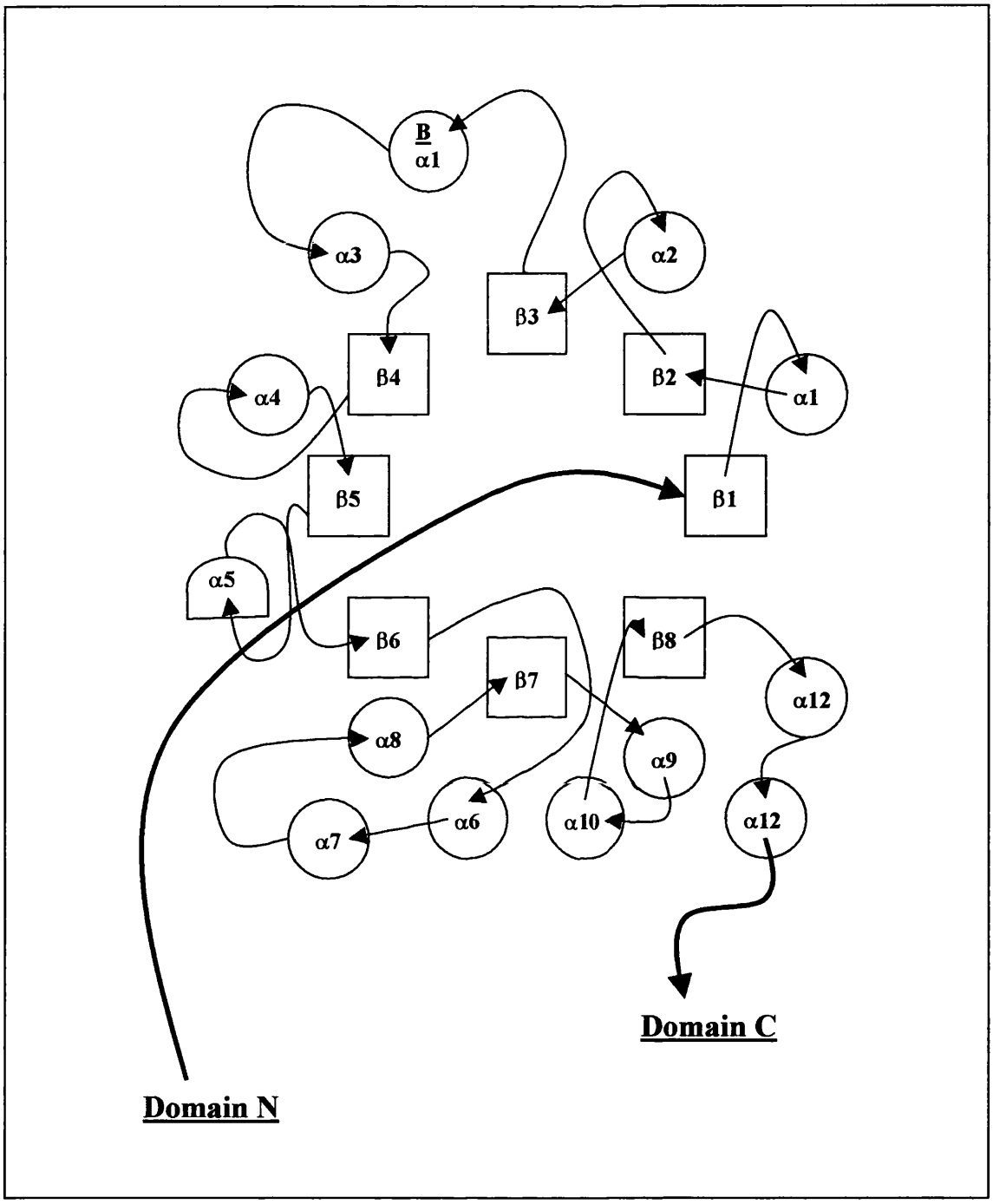


Fig.5.2: Topology of the α-helices and β-sheets that make up the Domain A (TIM barrel structure) of neopullulanase from *B.stearothermophilus*. Linking β3 to α3 is Domain B. The secondary structures of Domain N and C were not included. The N-terminus of the protein is found in Domain N and the C terminus is found in Domain C. The arrows in bold are loops that link Domain N to Domain A and Domain A to Domain C. This diagram is adapted from Hondoh *et al.*, 2003.

Table 5.5: Summary of the predicted secondary structures of Domain A in pullulanase from *T.natronophilum* based on collective data obtained by three different structure prediction approaches.

α/β	Predicted Secondary Structures in Domain A of Pullulanase from *T.natronophilum*
β1	^{340}IIYEI^{344}
β2	^{391}HVHL^{394}
β3	^{456}IRVILDMV^{463}
β4	^{529}GFRF^{532}
β5	^{557}VVLYGE^{562}
β6	^{584}GAF^{586}
β7	^{641}TINY^{644}
β8	^{692}VFL^{694}
α1	^{377}VTTGLDHLIE^{386}
α2	^{442}REVKQMIQALHE^{453}
α3	^{508}PMMRKYIIDTCKYWME^{523}
α4	^{541}VTMRTLAEE^{549}
α5	^{568}NAPVRFGKAHVG^{579}
α6/α6'	^{588}DDFRDAMR^{594}
α6/α7	^{618}RRGVA^{622}
α7	^{674}LRNAQKLAGAILL^{686}
α7'	^{653}LWDKNV^{658}
α8	^{731}VFEYYKGLIKLRRE^{744}

For α7, alignments with neopullulanase from *B.stearothermophilus* (1), indicates that ^{618}RRGVAVSIEYDERIR^{632} are the amino acids making up α7. ^{618}RRGVA^{622} were indicated to be part of α6 when aligned to data predicted by Kashiwabara *et.al.*, (1999) in Table 5.4. In PELE, all these amino acids were predicted to be α-helix residue. Thus, this stretch of sequence might be either α7 or α6 residues.

From data obtained from alignments with type I pullulanase of *K.aerogenes* W70, *B.stearothermophilus* TRS 128, *Thermus Sp.* AMD-33, *B.acidopullulyticus*, *C.saccharolyticus* and *T.maritima* (2), ^{674}LRNAQKLAGAILL^{686} was the predicted α7. From PELE, the first eight residues have been predicted to be α-helix residues. It is also possible to assume that this is the true α7 residues or perhaps an additional α-helix in the Domain A, in close proximity with the TIM barrel scaffolding.

An additional α-helix has been reported in isoamylase from *P.amyloderamosa* JD210 and oligo-1,6-glucosidase from *B.cereus* ATCC 7064 and this has been annotated as α7'. A homologous region to α7' has been identified in pullulanase peptide sequence from *T.natronophilum* as ^{653}LWDKNV^{658}. In PELE, these have also been identified as α-helix residues. It is very likely that this is indeed an extra helix in *T.natronophilum*.

From alignments with neopullulanase from *B.stearothermophilus* (1), ^{636}KNPQ^{639} were the α8 residues. From PELE, none of these residues were α-helix residue. In alignments with type I pullulanase of *K.aerogenes* W70, *B.stearothermophilus* TRS 128, *Thermus Sp.* AMD-33, *B.acidopullulyticus*, *C.saccharolyticus* and *T.maritima* (2), another stretch of amino acids identified as ^{731}VFEYYKGLIKLRRE^{744} were predicted to be α8 residues. In PELE, majority of the residues have been predicted to be α-helix residues.

These data remain predicted secondary structure analysis of Domain A of pullulanase from *T.natronophilum*.

5.7.3 Results and Discussions (Domain B)

According to data published by Hondoh *et al.* (2003), the loop linking β3 to α3 comprises residues 244 to 302 in the neopullulanase from *B.stearothermophilus*, although it is interrupted by a short helix ^{254}PFQDVM^{259}. Through computational analyses (Fig. 5.1), ^{481}PYYFYR^{486} of *T.natronophilum* was predicted to form a β-sheet. However, ^{472}PQSAFD^{477} in *T.natronophilum* is shown to align with ^{481}PYYFYR^{486} from *B.stearothermophilus* neopullulanase. Thus, although it is clear that residues 463 to 507 are part of the domain B of pullulanase from *T.natronophilum*, but the possible β-sheets or α-helices that are within this stretch of residues remains to be deciphered.

```
BSN_DomB   244VFNHCGYEFAPFQDVWKNGESSKYKDWFHIHEFPLQTEPRPNYDTFAFVPQMPKLNTAN302
Tn_DomB    463VFPHT-YGVGP-QSAFD--QTVPY--YFYRLD---KTGAYLNESGCGNVIASER-----507
              ** *  * ..* *..:.  ::  *  :*:  :   :* .  * .  . *    :
```

Fig. 5.3: Alignment of peptide sequences that make up Domain B in neopullulanase of *B.stearothermophilus* and pullulanase of *T.natronophilum*.

5.7.4: Results and Discussions (Domain C)

From alignments with neopullulanase from *B.stearothermophilus*, the predicted *β*-sheets are tabulated in Table 5.6. Data predicted from PELE is presented in Fig.5.4. By comparing both sets of data, most of the predicted sequences by alignment with *B.stearothermophilus* neopullulanase were also predicted in PELE. ^{760}LTF^{762}, ^{773}V, ^{734}ILVIY^{788}, ^{805}KV^{806}, ^{830}I and ^{834}VM^{835} (*T.natronophilum* numbering) are residues predicted in both approaches and most likely to be β-sheet residues in Domain C.

Table.5.6: Alignment of the domain C of pullulanase from *T.natronophilum* with neopullulanase from *B.stearothermophilus*.

	Neopullulanase *B.stearothermophilus*	Pullulanase *T.natronophilum*
β1	^{508}ISFLH^{512}	^{760}LTFLP^{764}
β2	^{520}LIKYYT^{525}	^{773}VLRDPK^{778}
β3	^{530}TVLVII^{535}	^{783}EILVIY^{788}
β4	^{540}QKADI^{544}	^{793}REQEF^{797}
β5	^{554}WLV^{556}	^{804}WKV^{806}
β6	^{563}RFA^{565}	^{812}RAG^{814}
β7	^{574}SL^{475}	^{823}GR^{824}
β8	^{581}VLYAIEHW^{588}	^{830}ISAMVMYR^{837}

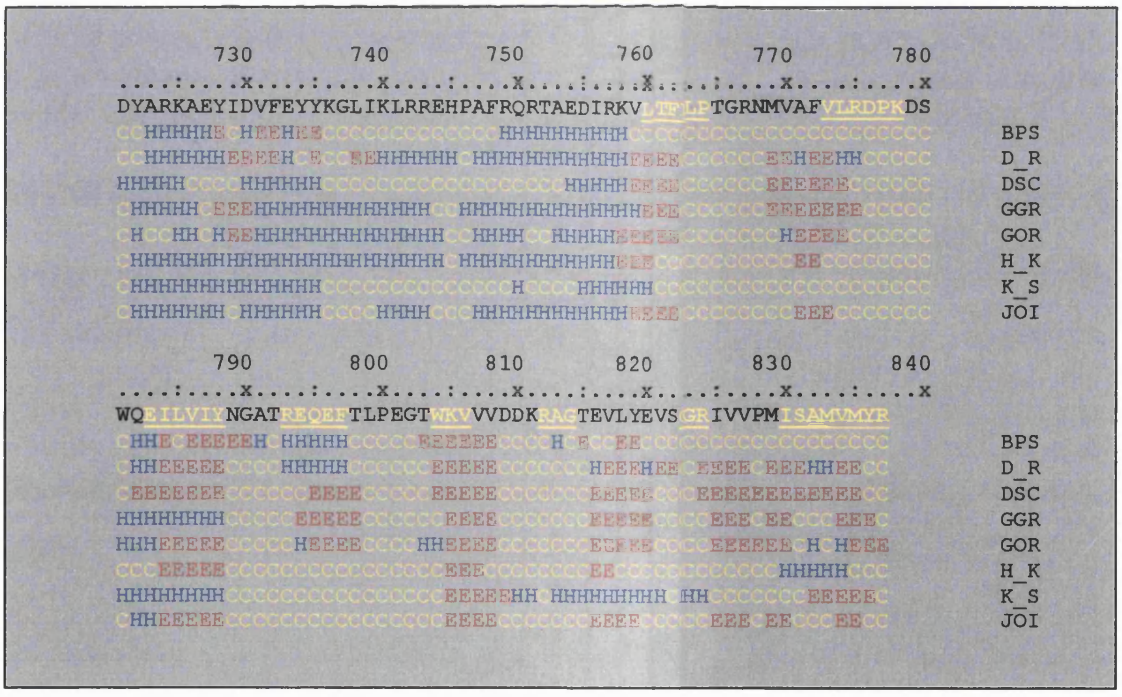


Fig.5.4: Predicted β-sheets in Domain C of *T.natronophilum* using PELE. The protein sequence above is protein sequence from *Thermopallium* pullulanase. Sequences highlighted in yellow and underlined are the predicted β residues in Table 5.6.

Section 4: Identification of Catalytic Residues of Pullulanase from *T.natronophilum*

5.8 Methods

5.8.1 Identifying Catalytic Residues

This is carried out by comparing the four highly conserved regions found in the amylase 13 family and the additional highly conserved region (annotated as region V in chapter 1) in most pullulanase type I enzymes. For the first four regions, cyclodextrin glycosyltransferase (CGTase) in *B.circulans* (Veen *et al.*, 2000) was used as the model of comparison. Type I pullulanase of *K.aerogenes* W70 (Yamashita *et al.*, 1997) was used as comparison in the identification of catalytic residues in Region V. Mutational studies were carried out on these two enzymes in order to identify residues important for catalytic activity or substrate binding.

5.9 Results

5.9.1 Residues Involved in the Double Displacement Mechanism

Even though members of the α-amylase family have less than 30% sequence identity, there are still some highly conserved residues involved in catalysis and substrate binding (Nakamura *et al.*, 1993). There are seven conserved residues with functions in total, which congregate together at the bottom of the substrate-binding groove at the surface of the TIM barrel (Uitdehaag *et al.*, 2002).

It has been established that two carboxylic acid groups, namely one glutamic acid and one aspartic acid, are crucial for the double displacement mechanism of substrate hydrolysis. In TAKA amylase of *A.oryzae*, these have been identified as Asp206 and Glu230 (Matsuura *et al.*, 1984). In CGTase of *B.circulans* strain 251, these have been identified as Asp229 and Glu257 respectively (Table 5.7) (Veen *et al.*, 2000). In type I pullulanase of *K.aerogenes* W70, these two carboxylic acids have been identified as Asp677 and Glu706 (Table 5.7).

Table 5.7: The sequence alignments of the conserved regions of CGTase from *B.circulans* strain 251 (CGT_BC), pullulanase type I from *K.aerogenes* W70 (PUL_KA), and pullulanase type I from *T.natronophilum* (PUL_TN).

	RI	RII	RIII	RIV	RV
CGT_BC	^{135}DFAPNH^{140}	^{224}DGIRMDAVKH^{233}	^{257}EWFL^{260}	^{323}FIDNHD^{328}	-
PUL_KA	^{602}DVVYNH^{607}	^{672}DGFRFDLMGY^{681}	^{706}EGWD^{709}	^{829}YVSKHD^{834}	^{555}YNWGYDP^{561}
PUL_TN	^{461}DMVFPH^{466}	^{528}DGFRFDQMGL^{537}	^{562}EPWG^{565}	^{644}YVASHD^{649}	^{414}YNWGYDP^{420}

As described in Chapter one, hydrolysis occurs by a double displacement mechanism (see Fig. 1.10). The equivalent of Asp229 in CGTase of *B.circulans* strain 251 has been identified as Asp533 in pullulanase from *T.natronophilum* (Table 5.7). This is the proposed nucleophile in all members of the α-amylase family. In the presence of a sugar, Asp229 is found to be located close to the glucose C1 atom and the *syn* lone pair orbital of this atom is positioned for nucleophilic attack by Asp229 (Strokopytov *et al.*, 1995; Brzozowski *et al.*, 1997).

Glu257 (CGTase of *B.circulans* strain 251 numbering) is the acid/base catalyst. In the absence of a ligand, Glu257 will form a hydrogen bond with Asp328 (Table 5.7) (CGTase of *B.circulans* strain 251 numbering). After a sugar binds to the active site, the side chain of Glu257 will form a hydrogen bond with the O4 atom in the scissile bond i.e. the point of cleavage. (Strokopytov *et al.*, 1995), in the case of α-1,4 bond hydrolysis. For type I pullulanase of *T.natronophilum*, Glu562 will probably form a hydrogen bond with O6, forming the oxo-carbonium transition state. This transition state is then stabilised by forming into an intermediate covalently linked to a nucleophile, identified as Asp229 in CGTase of *B.circulans* strain 251 (Veen *et al.*, 2000).

5.9.2 Residues Involved in Substrate Binding

Residues found to be involved in substrate binding are Asp328, His327, Arg227 and His140 (Table 5.7) (CGTase of *B.circulans* strain 251 numbering). Asp328 has been stated to play a vital role in stabilising the oxo-carbonium transition state by interacting with OH2 and OH3 of the sugar (Brzozowski *et al.*, 1997). This residue has also been speculated to maintain the protonation state of the Glu257 (Strokopytov *et al.*, 1995). The equivalent of Asp328 of *B.circulans* strain 251 has been identified as Asp649 in *T.natronophilum*.

The two basic residues, His327 and Arg227 (Table 5.7) (CGTase of *B.circulans* strain 251 numbering), are bound to OH2. His 327 was found to cause a distortion of the sugar ring to a half chair conformation while Arg227 reduces the electro-negativity of OH2 which would otherwise inhibit the formation of the positively-charged transition state (Uitdehaag *et al.*, 2002). The equivalents of His327 and Arg 227 have been identified as His648 and Arg531 (Table 5.7) in *T.natronophilum*.

Another conserved residue is His140 (Table 5.7) in CGTase of *B.circulans* strain 251 (Veen *et al.*, 2000), which has been identified as His607 in type I pullulanase in *K.aerogenes* W70 and has been shown to be involved in substrate binding. Mutation of His607 to Ala in *K.aerogenes* W70 resulted in the inability of the enzyme to bind to pullulan and thus resulted in complete loss of activity. The binding capacity of the mutated enzyme was tested by applying an enzyme sample onto an α-cyclodextrin affinity column. It was found that His607→Ala mutant did not bind to the column. The equivalent residue has been identified as His466 in *T.natronophilum* (Yamashita *et al.*, 1997).

5.9.3 Residue Crucial for the Architecture of the Catalytic Site

The mystery of the role of the last conserved residue, Asp135 (Table 5.7) (CGTase of *B.circulans* strain 251 numbering), was elucidated by Leemhuis *et al.* (2003). D135A and D135N mutants were studied. In both mutants, interactions between Glu257-Asp328 and Asp229-Arg227 were lost.

Asp135 was found to be important for the conformation of the side chain of Glu257 (the acid/base catalyst). In the D135N mutant, Glu257 side chain was found to turn away from the scissile bond, while in the case of D135A mutant, it either moved away from or towards the scissile bond.

Asp135 also has an effect on the nucleophile Asp229. The D135A mutation has caused the Asp229 residue to have two possible conformations, with one moving towards and one moving away from the scissile bond. The D135N mutant on the other hand does not affect the Asp229 residue at all. In this mutant, Asp229 is found to bind to O1, O5 and O6 atoms in the sugar molecule rather than O6 alone binding to Asp229, which is the case for the wild type enzyme.

In conclusion, Asp135 does not interact directly with the substrate but it is absolutely crucial for the sound architecture of the catalytic site. In type I pullulanase of *T.natronophilum*, this residue has been identified as Asp461 (Table 5.7) and most probably plays the same role as the homologous Asp found in *B.circulans* strain 251.

5.9.4 Residues Conserved in Pullulanases

In pullulanase of *K.aerogenes* W70, mutational studies of His833 to Ala were carried out. The mutant was found to bind to an α-cyclodextrin affinity column but was completely inactivated. Thus, Yamashita *et al.*(1997) proposed that this residue is important in the hydrolysis of α-1,6 glycosidic linkages. The equivalent residue has been identified as His648 (Table 5.7) in pullulanase of *T.natronophilum*.

Two tyrosine residues in pullulanase of *K.aerogenes* W70 were mutated, Tyr559→Ala/Phe and Tyr564→Ala/Phe respectively (Yamashita *et al.*, 1997). It was found that enzymes with Tyr559→Ala/Phe were completely inactivated, while enzymes with Tyr564Ala/Phe showed a decrease in enzyme activity. Both Tyr559 mutations did not affect the binding of the enzymes to substrate. This suggests that both mutations did not affect the conformation of the active site; rather mutation of Tyr559 affects the hydrolysis of α-1,6 glycosidic linkages. Yamashita *et al.*(1997) proposed that both tyrosines are part of a newly determined consensus sequences i.e.

^{558}GYD×××Y^{564} (*K.aerogenes* W70 numbering). In pullulanase from *T.natronophilum*, the equivalent of Tyr559 has been identified as Tyr 418, and in place of Tyr564 is Phe420. The latter case is also observed in pullulanase from *F.pennavorans Ven*5, *T.maritima* and *Bacillus cereus* ATCC 14579. However, in *Thermus Sp.* IM6501, *T.thermophilus* HB8 and *Geobacillus thermoleovorans*, the tyrosine is replaced by leucine (Table 5.8).

The hydrophobicity of the three residues in question follows the order F>L>Y, with F being the most hydrophobic (see Abbreviations on page vi). The preliminary observation is that all three residues are hydrophobic. Perhaps, hydrophobicity to a certain extent is important for the catalytic activity of this enzyme. Also, a replacement of this tyrosine with the more hydrophilic alanine resulted in a decrease in enzymatic activity. This also supports the previous observation. Considering the mutation studies of Tyr564→Phe has caused a drop in activity, it would be interesting to see if mutational studies of Phe420 in *T.natronophilum* to a tyrosine will actually increase the activity of the enzyme.

From alignments of the twelve different strains in Table 5.2 and Table 5.8, the majority of the type I pullulanases observed have a strictly conserved region of residues composed of ^{414}YNWGYDP^{420}. An anomaly is observed with type I pullulanase from *B.flavocaldarius*. The enzyme from this strain shows more homology with neopullulanase from *B.stearothermophilus*. Type I pullulanase from *B.flavocaldarius* was found to have 43% identity with neopullulanase from *B.stearothermophilus* TRS40 (Kashiwabara *et al.*, 1999).

Despite the differences observed in this region, Y418 and N415 (*T.natronophilum* numbering) remain strictly conserved in all type I pullulanases. N415, a hydrophilic residue (highlighted in red in Table 5.8), remains strictly conserved in the enzymes found in Table 5.8. It would be interesting to carry out mutational studies on N415 to investigate its role in α-1,6 glycosidic linkages hydrolysis and substrate binding.

Residue D419 (*T.natronophilum* numbering), corresponding to the sixth residue highlighted in yellow in Region V in Table 5.8, is a hydrophilic residue in all the enzymes listed.

Albertson *et al.* (1997) have proposed that true pullulanases will contain the conserved sequence of YNWGY and neopullulanases will have an equivalent of the same stretch of sequence as NH(K/R)Y. The anomaly presented by type I pullulanase from *B.flavocaldarius* has proved that these rules does not strictly apply.

In conclusion, individual mutational studies should be carried out on residues YNWGY-P in order to investigate the role of each amino acid in α-1,6 bonds hydrolysis or substrate binding.

Table 5.8: The conserved region found in type I pullulanase, neopullulanase and amylopullulanases. The residues highlighted in black are the proposed strictly conserved tyrosines in *K.aerogenes* W70 (Yamashita *et al.*, 1997). The first tyrosine corresponds to Y559 and Y564 in *K.aerogenes* W70. * indicates residues that are strictly conserved in Type I pullulanase.

Organisms	Enzyme	Region V
K.aerogenes W70	Type I Pullulanase	YNWGYDPFHY
T.natronophilum	Type I Pullulanase	YNWGYDPNLF
T.maritima	Type I Pullulanase	YNWGYDPYLF
*F.pennavorans Ven*5	Type I Pullulanase	YNWGYDPYLF
Thermus Sp. IM6501	Type I Pullulanase	YNWGYNPLHL
T.thermophilus HB8	Type I Pullulanase	YNWGYNPLHL
Bacillus cereus ATCC 14579	Type I Pullulanase	YNWGYDPKNF
Geobacillus thermoleovorans	Type I Pullulanase	YNWGYNPLHL
B.flavocaldarius	Type I Pullulanase	-NHRYHTVDY
B.stearothermophilus	Neopullulanase	-NHKYDTADY
B.halodurans	Amylopullulanase	YNWGYDPHSY
Thermoanaerobacterium thermosulfurigenes	Amylopullulanase	-NWGDASLDL
		-*--*-----

Section 5: Identification of the Signal Peptide of Pullulanase from *T.natronophilum*

5.10 Methods

5.10.1 Identifying the Signal Peptide

The signal peptide for the pullulanase gene was identified using a web based programme known as SignalP Version 2.0 (www.cbs.dtu.dk/services/signalP). The methods used are Neural Networks (NN) and Hidden Markov Models (HMM) trained on Gram-negative bacteria.

5.11 Results

5.11.1 Identifying the Signal Peptide

A signal peptide is normally composed of a positively charged region, known as the n-region, then a hydrophobic region, known as the h-region and lastly a neutral but polar region defined as the c- region (Nielsen, 1997). Before cloning was initiated, the signal peptide for the pullulanase type I was determined by analysing the first 70 amino acid from the N terminus using SignalP Version 2.0 (Nielsen *et al.*, 1997). The result for neural networks (NN) is annotated in the graphs in Fig. 5.5., which identify the signal peptide, with one parameter identifying the exact cleavage site while the other two discriminate signal peptide sequence from non-signal peptide sequence (Nielsen *et al.*, 1997). In Fig. 5.5, C-score identifies the cleavage site, the S score discriminates signal peptide from non-signal peptide sequence and the Y-score combines both C-score and S-score by pinpointing the position where C-score is high and S-score changes from a high to a low value (www.cbs.dtu.dk/services/signalP).

According to the data obtained in Fig. 5.5, the signal peptide is identified as the first 18 amino acids at the N terminus, designated as MKRIFSVVLLLTVVFLFA. The

cleavage site is between amino acid 18 and 19, with the former being an Alanine (A) and the latter a Glutamine (Q). Based on data obtained from SignalP, for pullulanase type I from *F.pennavorans Ven*5, the signal peptide cleavage site was found to be between an Alanine (A) and a Glutamic acid (E) (Bertoldo *et al.*, 1999).

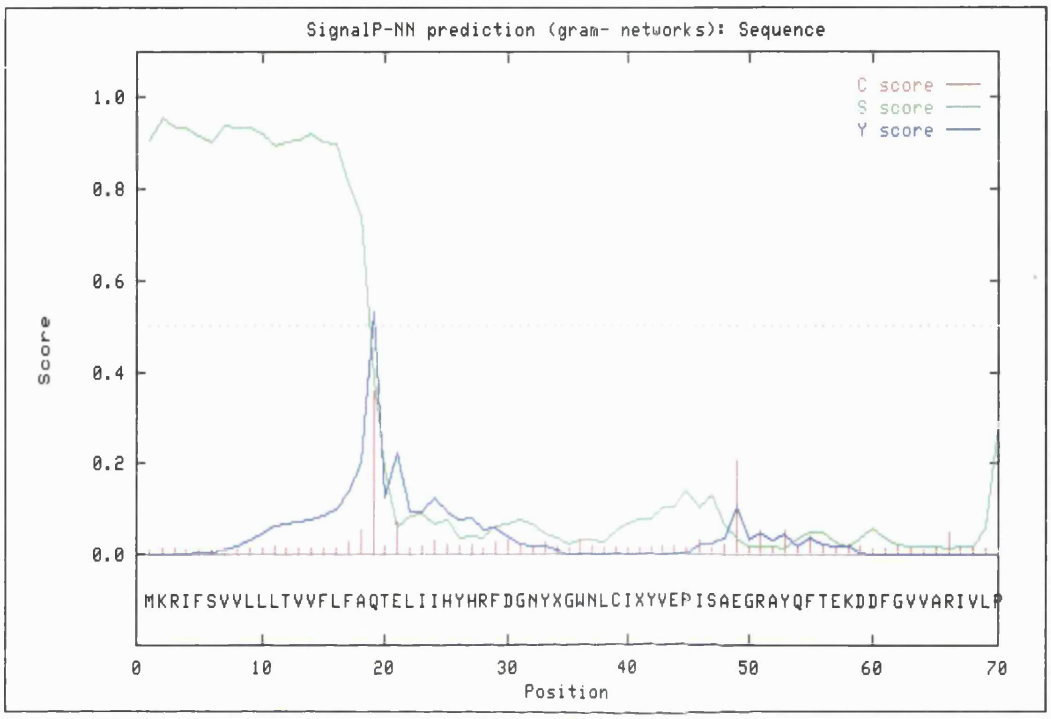


Fig. 5.5: The predicted cleavage site of the signal peptide of *Thermopallium* pullulanase type I using neural networks prediction.

The N terminus sequence of the mature pullulanase from *F.pennavorans Ven*5 has been determined and the predicted cleavage site was proved to be the true cleavage site of the signal peptide in this protein (Bertoldo *et al.*, 1999). When both pullulanase peptide sequences from *F.pennavorans Ven*5 and *T. natronophilum* were aligned, the cleavage sites of both enzymes were in perfect alignment and in conclusion the cleavage site predicted with SignalP in *T.natronophilum* is probably the true cleavage site in the native enzyme (Fig. 5.6).

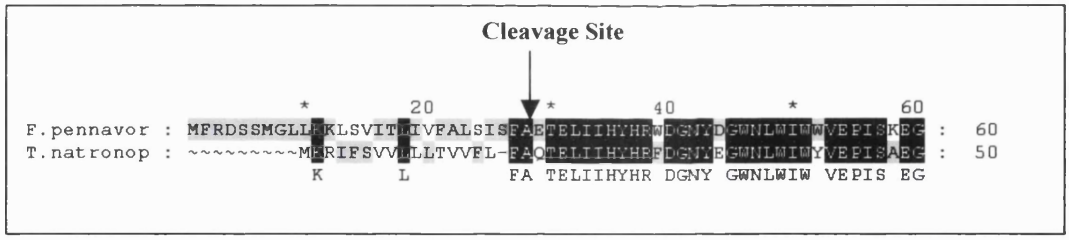


Fig. 5.6: The N terminus of both pullulanase type I from *F.pennavorans Ven5* and *T.natronophilum*. The arrow pinpoints the position of cleavage sites for both enzymes i.e. between A and E for *F.pennavorans Ven5*; between A and Q for *T.natronophilum*.

In the SignalP-HMM model, the parameters investigated rely on the nature of a signal peptide, whereby the initial part of a signal peptide is of a positively charged region, known as the n-region, then a hydrophobic region, known as the h-region and lastly a neutral but polar region defined as the c- region. According to the analysis obtained in Fig. 5.7, the signal peptide of the *Thermopallium* pullulanase is in the order of n-region, h-region and c-region. This is in good agreement with characteristics found normally in a signal peptide. Furthermore, both NN and HMM-based predictions identify the same cleavage site.

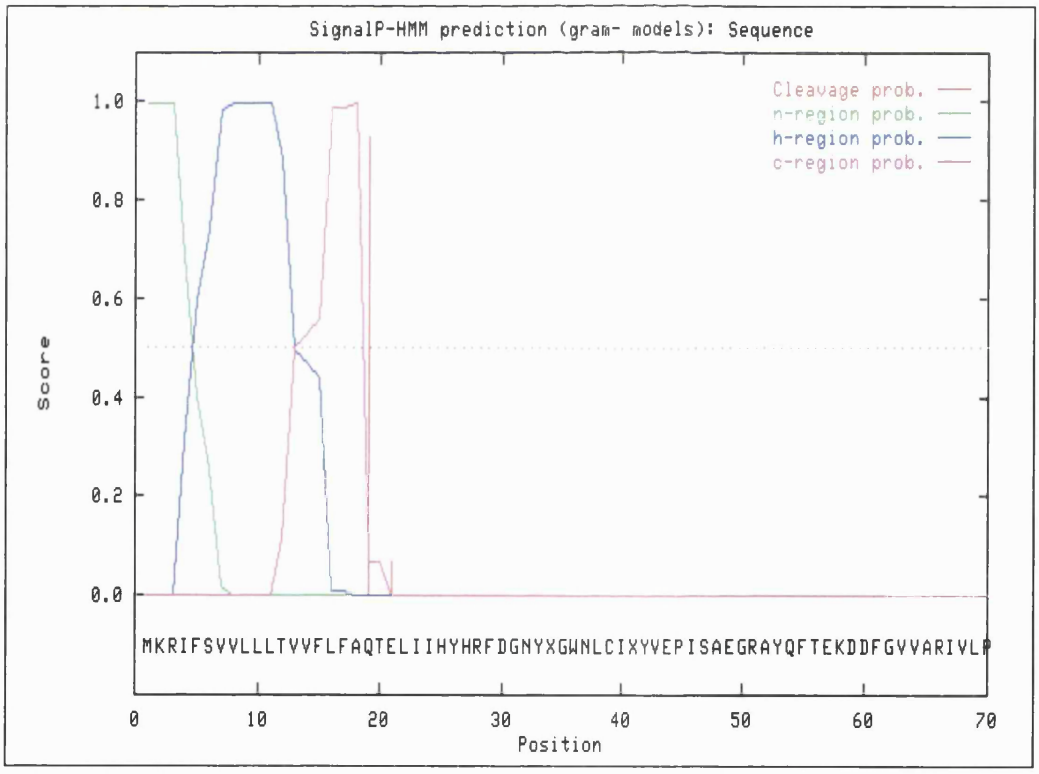


Fig. 5.7: The results obtained from SignalP-HMM model. The n-region denotes the positively charged region, the h-region denotes the hydrophobic region and the c-region denotes the neutral but polar region of the signal peptide.

Section 6: Identification of Potential Internal Start Site of Pullulanase Gene Sequence and Potential Protease VII Cleavage Site of Pullulanase Protein Sequence of *T.natronophilum*

5.12 Alternative Internal Start Site of Pullulanase Gene Sequence

After analysing the gene sequence of type I pullulanase of *T.natronophilum*, two possible internal start sites were identified. Earlier in Chapter 4, the possible –10 and –35 promoter sequences which will give a full length mRNA of approximately 2500kb were identified.

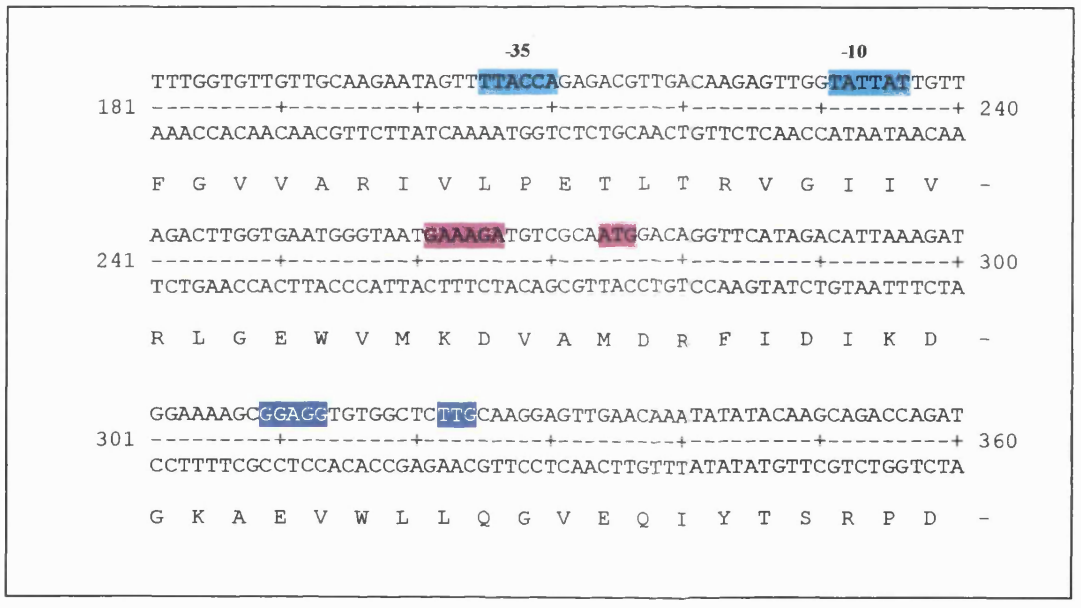


Fig. 5.8: The sequences highlighted in turquoise depict the potential –10 and –35 promoter sequence for transcription. The sequences highlighted in purple and blue the potential Shine-Dalgarno like sequence and ATG/TTG start sites for translation.

Here, an internal potential –10 and –35 promoter sequences were identified and has been highlighted as turquoise in Fig. 5.8. If transcription is also to take place at this position, then there are two internal start sites, each with their own potential Shine-Dalgarno like sequences. These two start sites have either ATG or TTG as the start site for translation. Translation may also begin at these two potential start sites from

the full length mRNA generated from transcription beginning from promoter sequences identified in Chapter 4. The truncated protein translated from the start site highlighted in purple (Fig. 5.8) will give an estimated Mr of 78kDa and the truncated protein translated from the start site highlighted in blue (Fig. 5.8) will give an estimated Mr of 76kDa. The latter start site has also been described in type I pullulanase from *F.pennavorans*.

5.13 Potential Protease VII Cleavage Sites on Type I Pullulanase Protein Sequence

Protease VII from *E.coli* is an endopeptidase that specifically cleaves between paired basic resides. It has a high specificity for Arg-Arg, followed by Lys-Arg and lastly Lys-Lys residue pairs (Sugimira and Nishihara, 1988). It has been found to have a higher preference for paired basic residues located near the C terminus. There are seven paired basic residues close to the C terminus of the pullulanase protein sequence (Fig. 5.9). Proteolysis between the paired basic resides will generated truncated proteins of estimated sizes of 54kDa, 59kDa, 66kDa, 78kDa, 80kDa, 87kDa providing the start site of the protein is right after the signal peptide of the protein.

```
501 NVIASERPMM RHYIIDTCKY WMEEYRVDGF RFDQMGLIDE VTMRTLAEEL
551 RKIDPAVVLY GEPWGGFNAP VRFGKAHVGG TGIGAFNDDF RDAMRGSVFN
601 PTVRGFLMGA LARETAIRRG VAGSIEYDER IRGFAKNPQE TINYVASHDN
651 HTLWDKNVLA AQADTRTQWT EEMLRNAQKL AGAILLTSQG PVFLHGGQDF
701 ARTKNFNENS YNAPISVNGF DYARKAEYID VFEYYKGLIK LRREHPAFRQ
751 RTAEDIRKVL TFLPTGRNMV AFVLRDPKDS WQEILVIYNG ATREQEFTLP
801 EGTWKVVVDD KRAGTEVLYE VSGRIVVPMI SAMVMYR
```

Fig. 5.9: The paired basic residues, which are potential cleavage points for protease VII, are all highlighted in black in the above pullulanase sequence.

Section 7: Identification of Cysteine Residues in the Type I Pullulanase Protein Sequence of *T.natronophilum*

5.14 Possibility of Disulphide Bridges in Type I Pullulanase Protein Sequence of *T.natronophilum*

The oxidation of two cysteine residues will result in the formation of a disulphide bond via covalent interaction. In proteins, disulphide bonds can be an inter or intra molecular bridges. Introduction of disulphide bridges by site directed mutagenesis has yielded mixed results, whereby some proteins are less stable than wild type (e.g. T4 lysozyme) while some have shown greater stability than wild type (e.g. RNase Hn) (http://www.cryst.bbk.ac.uk/PPS2/projects/day/TDayDiss/DisulphideBonds.html). In hyper-thermostable protein, disulphide bonds are rarely found. This is because these covalent bonds are heat labile.

By analysing the protein sequence, five cysteine residues have been identified in type I pullulanase sequence of *T.natronophilum* (Fig.5.10).

```
351 GLPNSGVKNR ATYLGLTERG TRGPNGVTTG LDHLIELGIT HVHLLPIFDF
401 ATCDETCRDF EKCYNWGYDP NLFTVPEGRY ATDPYDPYVR IREVKQMIQA
451 LHENGIRVIL DMVFPHTYGV GPQSAFDQTV PYYFYRLDKT GAYLNESGCG
501 NVIASERPMM RKYIIDTCKY WMEEYRVDGF RFDQMGLIDE VTMRTLAEEL
```

Fig. 5.10: Cysteine residues identified in the protein sequence of type I pullulanase from *T.natronophilum* are highlighted in turquoise.

Eight cysteine residues were identified in neopullulanase from *B.stearothermophilus* but no disulphide bonds was found in the protein (Kuriki and Imanaka, 1989). Thus, there is a highly possibility that only free thiols are present in the type I pullulanase of *T.natronophilum*, but further experiments need to be carried out to verify this.

Chapter 6: Cloning, Expression and Purification of Pullulanase Type I of *T.natronophilum*

6.1 Introduction

After the identification of the *T.natronophilum* pullulanase gene, the cloning of the gene into an expression vector was pursued. Having predicted the signal peptide of the pullulanase protein (Chapter 5), the pullulanase gene without the region encoding for the signal peptide was amplified using PCR. The amplified gene was first sub-cloned into a carrier vector (pGEM®-T Easy) before cloning it into an expression vector (pET28a).

Following the successful cloning of the pullulanase gene into an expression vector, the clone was then sequenced and retransformed into an expression host, namely *E.coli* BL21 (DE3) (Novagen, UK). Following this, expression of the recombinant pullulanase was monitored. Then, the recombinant protein was purified by affinity chromatography and anion exchange chromatography.

This chapter is divided into three sections: the cloning of the pullulanase gene, expression of the recombinant pullulanase, and the purification of the recombinant pullulanase.

6.2 Methods

6.2.1 Amplification of Pullulanase Gene without the Signal Peptide Sequence

PCR was carried out as described in section 2.1.6, using 2U of Vent$_R$® DNA Polymerase. Single primer controls were set up as described. Primers used in this PCR reaction are found in Fig. 6.1. Primers used have two annealing temperatures. The lower Tm is specific to the pullulanase gene while the higher Tm corresponds to the rest of the primer with the engineered restriction sites. PCR was carried out under the conditions found in Table 6.1.

Table 6.1: PCR conditions used in the amplification of the pullulanase gene.

Temp (°C)	Time (min)	No. of Cycles	Temp (°C)	Time (min)	No. of Cycles
96°C	3 min	1×	96°C	1 min	15×
96°C	1 min	15×	69°C	1 min	15×
58°C	1 min	15×	72°C	5 min	15×
72°C	5 min	15×	72°C	10 min	1×

CP/Fwd/*NcoI*

```
              M   A   Q   T   E   L   I   I   H   Y
5'-CAT ACC ATG GCT CAA ACT GAA CTT ATT ATC CAC TAC C-3'
          Degeneracy: 1     Tm: 60.1°C/71.8°C
```

CP/Rvr/*BamHI* (Reverse and Complement)

```
          V   M   Y   R   *   Q   K   L   S   V
(N) 5'-G GTG ATG TAC AGA TAA CAA AAA CTT AGT GTA GGA TCC GCG-3'
(C) 3'-C CAC TAC ATG TCT ATT GTT TTT GAA TCA CAT CCT AGG CGC-5'
(R) 5'-CGC GGA TCC TAC ACT AAG TTT TTG TTA TCT GTA CAT CAC C-3'
          Degeneracy: 1     Tm: 60.1°C/75.6°C
```

Fig.6.1: Primers used for the amplification of pullulanase gene. Both primers have restriction enzyme sites engineered into them (coloured blue). The forward primer had a *NcoI* restriction site and the reverse primer had a *BamHI* restriction site. Sequences prior to the restriction sites play a role in increasing the restriction efficiencies. * in the sequence represents the STOP codon. (N)=normal; (C)=complementary;(R)=Reverse and complementary. (R) is used for PCR.

6.2.2 A-Tailing of PCR Product with Taq Polymerase

The A-tailing reaction was at first set up according to the manufacturer's instructions but a scale-up was carried out and is described in Table 6.2. The Taq Polymerase used is from Promega, UK. The reaction was incubated at 70°C for 1 h.

Table 6.2: The components of the A-tailing reaction.

Reagents	Volume (µl)
Taq Polymerase (5U/µl)	2µl
10× Taq Buffer	2µl
25mM $MgCl_2$	2.4µl
1mM dATP	4µl
PCR Product (20ng/µl)	9.6µl
Final Volume	20µl

6.2.3 Ligation of A-Tailed Pullulanase Gene into pGEM®- T Easy Vector

For the ligation into pGEM®-T Easy vector, the A-tailed pullulanase gene can be used directly from the A-tailing reaction without any subsequent DNA purification steps. A positive ligation control was carried out by ligating the supplied control insert into the vector. Both reactions were set up as described in Table 6.3. The ligation reaction was carried out at 4°C overnight. Following this, the entire ligation reaction mix was transformed into JM109 heat shock competent cells, based on the method described in section 2.1.15.1, and plated out on LB/Amp plates.

Table 6.3: Ligation of the A-tailed pullulanase gene into carrier vector pGEM-T Easy vector.

	Standard Reaction	Positive Control
2x Rapid Ligation Buffer	10µl	5µl
pGEM®-T Easy Vector (50ng/µl)	1µl	1µl
PCR Product	7µl	-
Control Insert DNA	-	2µl
T4 DNA Ligase (3U/µl)	2µl	1µl
MQH_2O	-	1µl

6.2.4 Restriction Digest of pGEM-T/Pullulanase Clone with *Nco*I and *BamH*I and *Bsa*I to Prepare the Insert stock

The pGEM®-T Easy /pullulanase clone was first grown in a 10ml LB/Amp culture at 37°C overnight. The cells were then harvested and a plasmid prep was carried out, using the Nucleospin® Plasmid Kit (BD Biosciences Clonetech, Ca, USA) as per the manufacturer's instructions. 20µl of plasmid DNA was first digested at 37°C overnight with 10U of *Nco*I (NEB, UK) in 10× NEB buffer 4 (a final concentration of 1× was used) in a final volume of 50µl. Following the digestion of plasmid with *Nco*I, the linearised plasmid was gel purified using QiaexII Gel Purification Kit (Qiagen, Germany). The plasmid DNA was eluted with 20µl of MQH_2O. Subsequently, the linearised plasmid DNA was digested with 10U of *BamH*I and 10U of *Bsa*I in 1× *BamH*I buffer and 1× BSA in a final volume of 50µl. The sample was then incubated at 37°C for 2 h before being gel purified using QiaexII Gel Purification Kit (Qiagen, Germany). The double-digested plasmid was resuspended in 20µl of MQH_2O.

6.2.5 Preparation of Expression Vector pET28a

1-3μg of pET28a were digested with *Nco*I as described in section 6.2.4 and gel purified using QiaexII Gel Purification Kit (Qiagen, Germany). Subsequently, the *Nco*I linearised pET28a was digested with 20U *BamH*I and dephosphorylated with 2U of Shimp Alkaline Phosphatase (Roche, Germany). Both enzymatic reactions were carried out in 1× *BamH*I buffer and 1× BSA in a final volume of 50μl for 3 h at 37°C. The whole sample was then gel purified using QiaexII Gel Purification Kit (Qiagen, Germany).

6.2.6 Ligation of Insert with Expression Vector

Ligation was carried out using T4 DNA Ligase from Promega, UK. Before ligation was carried out, the concentration of insert and vector was determined by running 2μl of each on a 0.7% agarose gel alongside High Molecular Mass standards (Invitrogen, UK). Ligation was carried out at 4°C overnight and the ligation reaction and its negative control were set up as described in Table 6.4. Following this, 2μl of the ligation mixture was then transformed into JM109 heat shock competent cells, plated on LB/Kan plates and incubated overnight at 37°C.

Table 6.4: Ligation of insert with expression vector pET28a.

	pET28a/*NcoI*/*BamHI*/ Dephosphorylated and Insert/*NcoI*/*BamHI*	**pET28a/*NcoI*/*BamHI*/ Dephosphorylated**
pET28a/*NcoI*/*BamHI*/ Dephosphorylated (50ng/μl)	2μl	2μl
Insert/*NcoI*/*BamHI* (50ng/μl)	6μl	-
10× Ligase Buffer	1μl	1μl
T4 DNA Ligase (400U/μl)	1μl	1μl
MQH_2O	-	6μl
Final Volume	10μl	10μl

6.2.7 Restriction Digest of pET28a/Pullulanase Clone with *Nco*I and *BamH*I to Ensure the Cloning was Successful

10 colonies were picked and grown in 10ml of LB supplied with 100μg/ml kanamycin overnight at 37°C. Plasmids of 10 clones were extracted using the Nucleospin® Plasmid Kit (BD Biosciences Clonetech, Ca, USA) as per the manufacturer's instructions. 20μl of each plasmid was then double digested with 10U *Nco*I and 10U *BamH*I in 1× *BamH*I buffer and 1× BSA in a final volume of 60μl reaction. Reactions were incubated overnight at 37°C and 8μl of each digested material were then run on a 0.7% agarose gel.

6.3 Results

6.3.1 Cloning of Pullulanase Gene into pGEM-T® Easy Vector

Following the identification of the signal peptide, specific primers (Fig. 6.1) were designed to amplify the pullulanase gene sequence. The forward primer has been designed to exclude the signal peptide, and this therefore necessitated the engineering of a methionine codon and an *Nco*I restriction site into the sequence. To ensure that the primer contains the *Nco*I recognition site while maintaining the correct reading frame, two additional nucleotides are needed immediately after the *Nco*I restriction site (Fig. 6.1). Consequently, the Alanine before the signal peptide cleavage site of the *Thermopallium* pullulanase was also incorporated into the forward primer.

A *BamH*I restriction site was engineered into the reverse primer. The stop codon for the pullulanase gene, followed by 15bp that are not part of the pullulanase gene, were also incorporated into the primer. For the amplification of the pullulanase gene, single primer controls were also carried out. PCR samples were run on a 0.7% agarose gel and the results obtained are shown in Fig. 6.2.

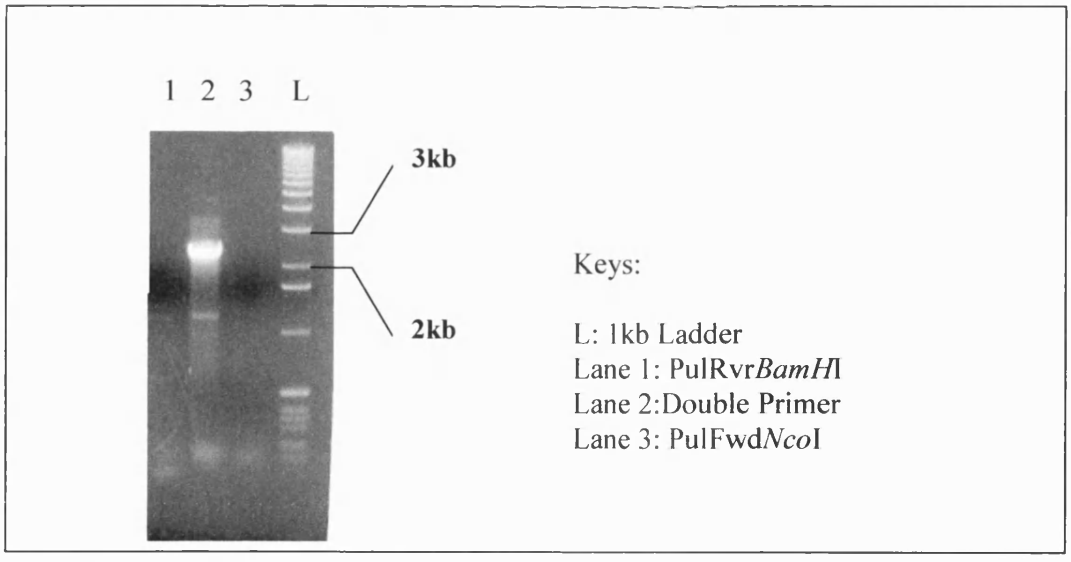


Fig. 6.2: Agarose-gel electrophoresis of the PCR product of 2.5kb band (Lane 2) amplified from PulRvr*BamHI* and PulFwd*NcoI*, using gDNA as a template.

The DNA band of approximately 2.5kb was excised and gel purified. Approximately 200ng of the PCR product amplified was then A-tailed with Taq polymerase in the presence of 1mM dATP as described in section 6.2.2. Following this, the A-tailed pullulanase gene was then ligated with 50ng of pGEM-T® Easy Vector as described in section 6.2.3. pGEM-T® Easy Vector allows for blue-white selection, and several white colonies were obtained from this experiment. One of the clones was picked and double digested with *Nco*I and *BamH*I and this resulted in the production of 2 bands, which correspond to the 3kb vector and a 2.5kb insert (Fig. 6.3). From these preliminary studies, the clone was deduced to be positive and thus sent for sequencing. The sequence of the clone was compared with the original gene sequence obtained and was found to be identical.

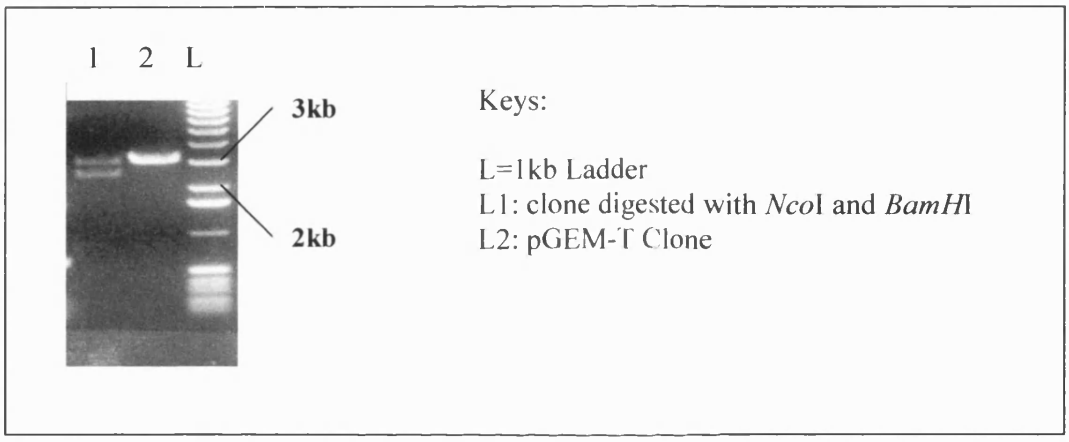


Fig. 6.3: Agarose-gel electrophoresis of PCR products. In Lane 1, the 3kb fragment corresponds to the pGEM®-T Easy vector while the 2.5kb corresponds to the pullulanase gene. Lane 2 is the undigested clone.

6.3.2 Cloning of the Pullulanase Gene into Expression Vector pET28a

With the pullulanase gene successfully cloned into the pGEM®-T Easy vector, a stock of insert was then prepared. Initially, the pGEM®-T Easy vector clone was digested with *Nco*I and *BamH*I, resulting in the formation of 2 bands, i.e., the 2.5kb insert and the 3kb pGEM®-T Easy vector. The 2.5kb insert was gel purified. The undigested pGEM®-T Easy vector clone runs at the same level as the 3kb fragment in a 0.7% agarose gel (Fig. 6.3). There is only a 500bp difference between vector and insert,

which initially created some false positives in the cloning process because of the difficulty in getting clean excision of the correct band from the DNA gel.

To overcome this, the pGEM®-T Easy vector clones were first digested with *Nco*I and the 5.5kb linearised plasmid generated was gel purified and then double digested with *BamH*I and *Bsa*I. *Bsa*I will digest pGEM®-T Easy vector to produce 2 bands of 1471bp and 1544bp (Fig. 6.4).

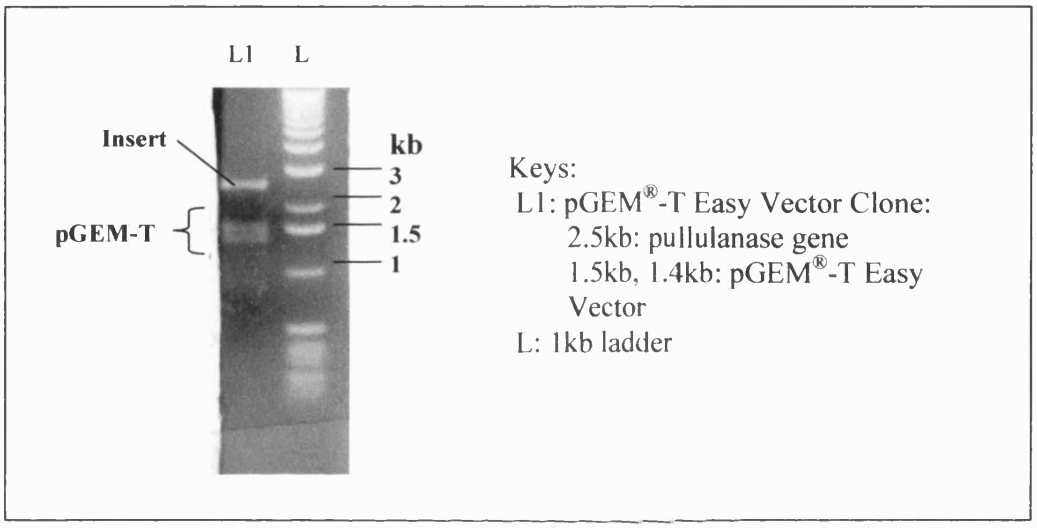


Fig. 6.4: Agarose-gel electrophoresis after digestion of pGEM®-T Easy vector clone with *Nco*I, *BamH*I and *Bsa*I to make insert stock.

pET28a was chosen as the expression vector because it employs kanamycin as the selection marker. By using this approach, any contaminating pGEM®-T Easy vector (which uses ampicillin as the selection marker) in the insert stock even though transformed will not be able to survive on a kanamycin LB agar plate. Ligation was carried out under a vector: insert ratio of 1:5. To check whether the clones were positive, 10 colonies were picked and plasmid extracted and double digested with *Nco*I and *BamH*I. Double digestion resulted in the production of 2 fragments of sizes 5kb and 2.5kb, which corresponds to the linearised pET28a vector and pullulanase gene insert (Fig.6.5). The result obtained was in good agreement with the calculated fragments size.

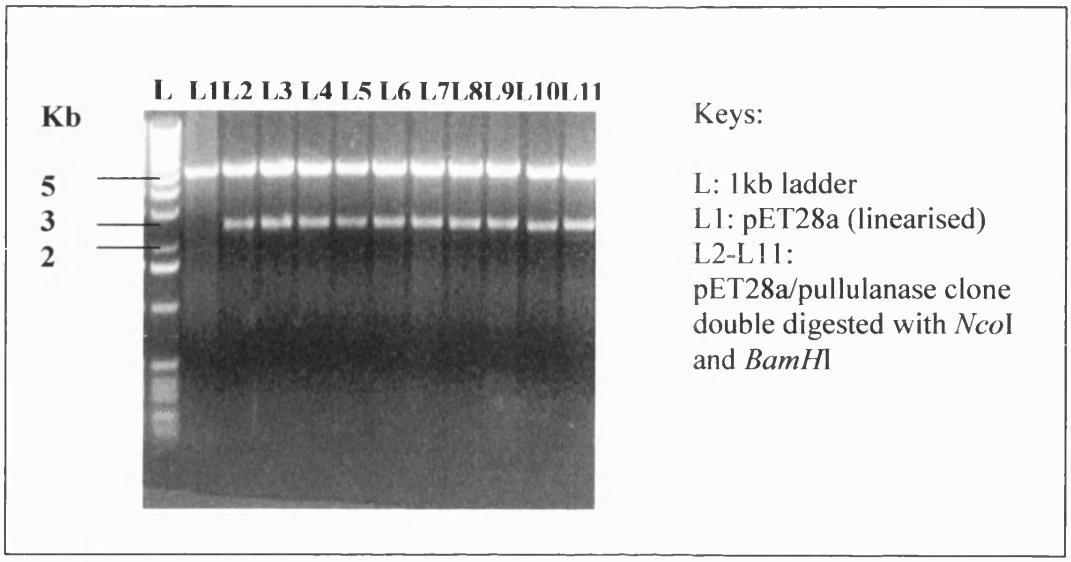


Fig.6.5: Agarose-gel electrophoresis after double digestion of pET28a/pullulanase clone with *Nco*I and *BamH*I. L1 is the linearised pET28a vector (control).

After the successful cloning of the pullulanase gene into an expression vector, the plasmid was transformed into the expression host *E.coli* BL21 (DE3).

Section 2: Expression of Recombinant Pullulanase

6.4 Methods

6.4.1 Transformation of pET28a/pullulanase into Expression Host BL21 (DE3)

All 1.5ml tubes were pre-chilled on ice. 50ng of pET28a/pullulanase vector was mixed with 50µl of *E.coli* BL21 (DE3) competent cells with *lac UV*5 promoter and left on ice for 30 min. Following this, the 1.5ml tubes were then heated at 42°C for 20 sec. The tube was then transferred back on ice for 2 min. Following this, 37°C preheated SOC medium (0.94ml) was added to the tube and the tube was then shaken at 170rpm for 1h before plating 100µl on LB agar/Kanamycin (0.1mg/ml) plates. These plates were then incubated at 37°C overnight. 50ng of pET28a vector was also transformed. This was the negative control for recombinant protein expression.

6.4.2 Expression of Recombinant Pullulanase by Induction with IPTG

One colony of the *E.coli* BL21 (DE3)/pET28a/pullulanase clone was picked and grown in 10ml of LB medium supplied with 30µg/ml of Kanamycin in a 50ml Falcon tube. One colony of *E.coli* BL21 (DE3)/pET28a was also grown. Both cultures were shaken at 250 rpm at 37°C till an OD_{600} of 0.6 was reached. Cells were then spun down and resuspended with 10ml of fresh LB and 30µg/ml of Kanamycin. 2ml of this inoculum were then used to inoculate 50ml of LB with 30µg/ml of Kanamycin in a 250ml flask and this was then grown at 37°C at 200 rpm till an OD_{600} of 0.7 was reached. Following this, IPTG was added to a final concentration of 1mM for protein induction. Before induction, a 1ml sample was kept as a control. Once induced, the cells were grown at 37°C and 200 rpm and two 1ml samples were taken at an hourly time point for 4 h. One set of the 1ml sample was to be assayed for activity. Each sample was sonicated at 3×15 s bursts at 16-18microns, (3mm probe, 150W Ultrasonic Disintegrator Mk 2, MSE, Crawley, UK) and the supernatant recovered by removing cell debris by centrifugation at 13,000 ×g for 3 min at 4°C. 100µl of the supernatant was then used in the standard assay for pullulanase activity.

The rest of the 1ml samples were spun at 13,000 ×g and 3 min to pellet the cells. The supernatant was removed and the pellet was then resuspended in 100µl of BugBuster® Protein Extraction Reagent (Novagen, UK). Samples were left standing at room temperature for 30 min before being sonicated at 3×10 s bursts at 16-18microns, (3mm probe, 150W Ultrasonic Disintegrator Mk 2, MSE, Crawley, UK). Cellular debris was removed by centrifugation at 13,000 ×g for 3 min at 4°C. 15µl of the supernatant was then mixed with 15µl of 2× SDS loading dye. This was then run on a 10% SDS PAGE.

Following induction for 4 h, the culture of *E.coli* BL21 (DE3)/pET28a/pullulanase clones were harvested by centrifuging cells at 5,000 ×g for 20 min at 4°C. The pellet obtained was then resuspended in 10ml of 50mM Tris-HCl, pH 8.5, 5mM EDTA. Cells were then sonicated at 3× 30 s bursts at 24microns. The cell suspension was then aliquoted to 1.5ml tubes and cell debris was then spun down at 13,000 ×g for 5 min at 4°C. The supernatant from these were then pooled together. The cell extracts can now be used for purification.

6.5 Results

E.coli BL21 (DE3) transformed with pET28a vector alone was used as the negative control for recombinant pullulanase expression. It is recommended in the pET Manual System 10th Edition (Novagen) that induced cells should be harvested 2-3 h after induction in order to prevent the culture from overgrowing, thus causing the possibility of loosing the plasmid (pET System Manual, 10th Edition, Novagen UK). Thus, to check the level of expression of the recombinant pullulanase, 1ml samples were taken at hourly time points after induction. Samples were collected till the 4th h.

Table 6.5: Protein expression before and after induction with IPTG.

Events	Enzyme Activity (U)
Before Induction	0.038
1 h after Induction	0.045
2 h after Induction	0.048
3 h after Induction	0.056
4 h after Induction	0.045

Table 6.5 shows the pullulanase expression level before and after induction with IPTG. Pullulanase activity detected before induction with IPTG reflects that the host expression strain is leaky.

Samples from each hourly time point were also compared by 10% SDS-PAGE. By comparing lane 2 to lane lane 6 in Fig. 6.6, the expression level increase at the hourly time point can be observed. Thus, based on the activity measurements, it appears that maximal protein expression can be achieved after 3 h at 37°C.

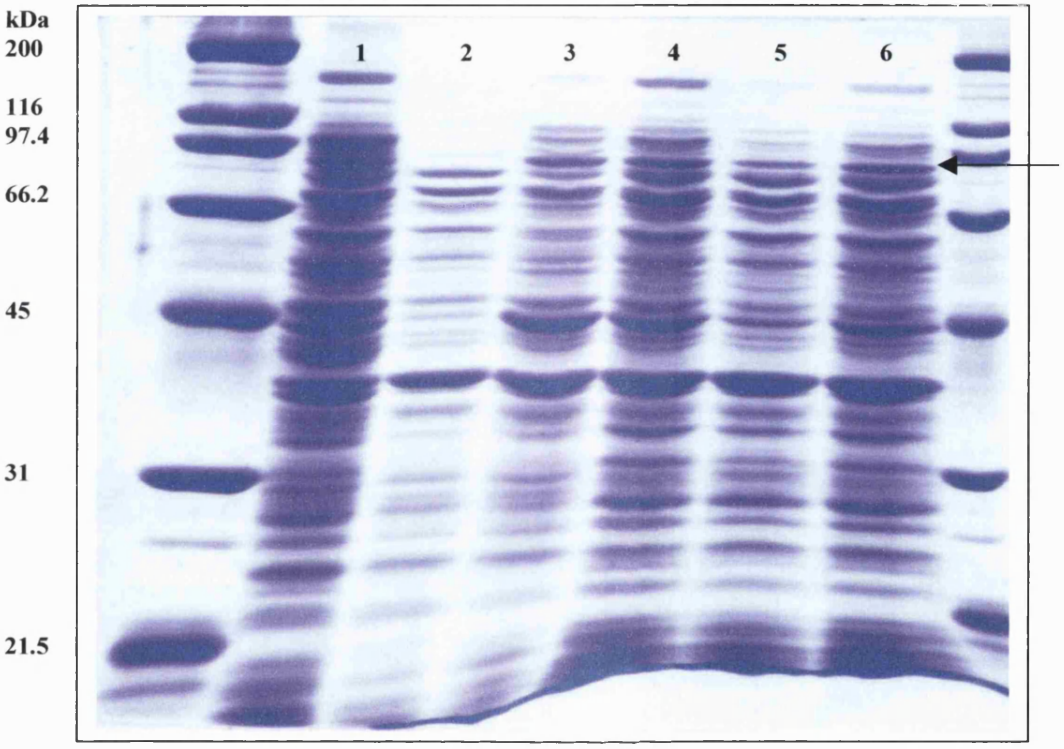


Fig. 6.6: SDS PAGE showing the expression of recombinant pullulanase at an hourly time point up to 4h.
Lane 1: BL21 (DE3)/pET28a (Control);
Lane 2: BL21 (DE3)/pET28a/Pullulanase Before Induction;
Lane 3: 1h After Induction;
Lane 4: 2h After Induction;
Lane 5: 3h After Induction;
Lane 6: 4h After Induction
The arrow indicates the predicted band corresponding to the recombinant pullulanase.

Section 3: Purification of Recombinant Pullulanase

6.6 Methods

6.6.1 Preparation of Affinity Chromatography Matrix

5g of freeze-dried epoxy-activated Sepharose 6B (Sigma Aldrich, Dorset, UK) were reconstituted with 10ml of MQH_2O overnight till the Sepharose was swollen; the excess MQH_2O was then decanted. The Sepharose was then resuspended in 7ml of 0.1M NaOH, and excess liquid was removed.

4g of α-cyclodextrin were dissolved in 12ml of 0.1M NaOH. Then, both the epoxy-activated Sepharose 6B and the α-cyclodextrin were mixed together and incubated at 45°C, 150 rpm for 16 h. The matrix was then washed with 50ml of MQH_2O for 30 min , 50ml of 2.5%(w/v) glucose for 30 min and finally with 50ml of MQH_2O for 30 min. All washes were carried out on a glass filter. Prior to packing into a column, the matrix was washed with 50ml of 20mM Tris-HCl, pH 8.0 for 2 h. The matrix was packed into a 20ml column to a matrix height of 8cm.

6.6.2 First Stage Purification of Recombinant Pullulanase Using Affinity Chromatography

The α-cyclodextrin-Sepharose column was connected to a Fast Protein Liquid Chromatography (FPLC) system supplied by Pharmacia, Uppsala, Sweden. The column was first washed with three column volumes of 20mM Tris-HCl, pH 8.5. Following this, 2ml cell extract was loaded and injected into the column at a flow rate of 1ml/min. During the absorption process, the column was continuously washed with 20mM Tris-HCl, pH 8.5. Protein that did not bind to the column was detected at 280nm. When the A_{280} had once again reached the base-line level, the recombinant pullulanase was then eluted off the column by washing the column in 20mM Tris-HCl, pH8.5, 0.2% α-cyclodextrin and collected as a 1ml fraction. During the elution process, a smaller peak was observed and fractions from this peak were then assayed for activity. Following this, 10μl of each fraction was visualised by SDS PAGE to

determine the degree of purification. Selective fractions were then pooled for the next stage of purification.

6.6.3 Second Stage Purification of Recombinant Pullulanase Using Gel Filtration

Fractions pooled from the previous purification were subjected to gel filtration using a Superdex 200 H 16/60 column, which was connected to an FPLC system, at 1ml/min flow rate. The buffer used was Tris-HCl, pH 8.5 and 1ml fractions were collected. Fractions with the highest activity were selected and 10µl of this were visualised on the SDS PAGE to determine the degree of purification.

6.6.4 Third Stage Purification of Recombinant Pullulanase with Anion Exchange Chromatography

A Mono Q (H 5/5) anion exchange column was connected to the FPLC system and samples were loaded at a flow rate of 1ml/min. The column was washed with 20mM Tris-HCl, pH 8.5 and unbound fractions were collected and assayed for activity. Bound protein was eluted with 60ml of buffer with a salt gradient of 0 to 0.8M NaCl. Samples were collected as 1ml fractions and were assayed for activity. Fractions with the highest activity were also visualised on SDS PAGE to determine the degree of purification.

6.6.5 Determination of the MR of the Native and Recombinant Pullulanase

The polypeptide Mr values of the native and recombinant pullulanase were determined by measuring the distance travelled by each on 10% SDS PAGE using the following standard protein markers: myosin (200kDa), beta galactosidase (116kDa), BSA (97.4kDa), carbonic anhydrase (66.2kDa), soybean trypsin inhibitor (45kDa), lysozyme (31kDa) and aprolinin (21.5kDa).

6.7 Results

6.7.1 First Stage Purification of Recombinant Pullulanase Using Affinity Chromatography

Affinity chromatography has proven to be a very effective method in purifying recombinant pullulanase. Due to the small capacity of the column, overloading of the column is often observed. These unbound fractions were pooled and subjected to second round of affinity chromatography. In Fig. 6.7, by comparing the cell extract of *E.coli* BL21 (DE3)/pET28a/Pullulanase (Lane 1) and the eluted fractions (Lane 3-Lane 8), only two predominant protein bands were observed after affinity chromatography. The upper band most probably corresponds to the recombinant pullulanase. The lower band of approximately 66kDa was to be removed via purification and assayed for pullulanase activity. The second stage of purification was with gel filtration.

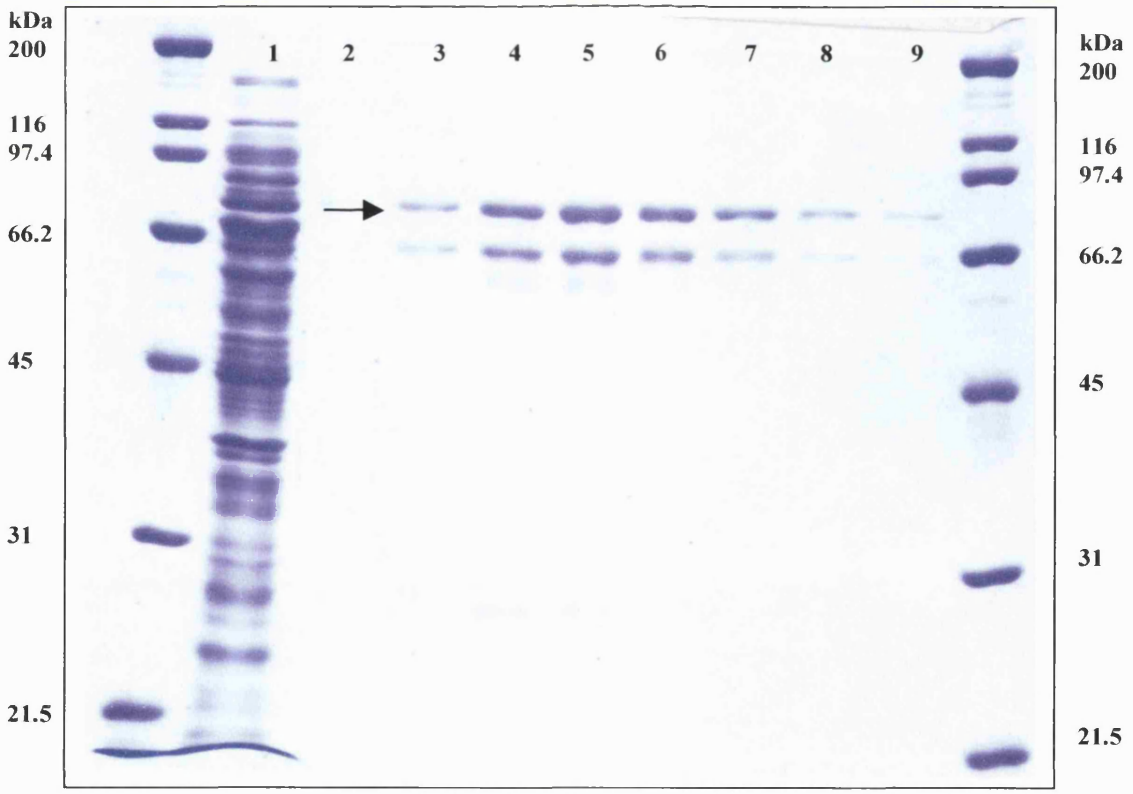


Fig. 6.7: SDS-PAGE of the eluted fractions of recombinant pullulanase from affinity chromatography. Lane 1: Cell extract: Lane 2-Lane 9: Fractions eluted with 0.2% α-cyclodextrin. The arrow indicates the predicted position of the recombinant pullulanase.

6.7.2 Second Stage Purification of Recombinant Pullulanase Using Gel Filtration

Gel filtration on H 16/60 Superdex 200 was not successful in separating the two predominant protein bands from affinity chromatography and this can be observed in Fig. 6.8. Lane 3, which is the sample from affinity chromatography, is under loaded in comparison to the gel filtration sample in Lane 4. This gel shows that gel filtration did not resolve these two protein bands and so anion exchange chromatography was chosen to be the next approach to separate the recombinant pullulanase from contaminating proteins.

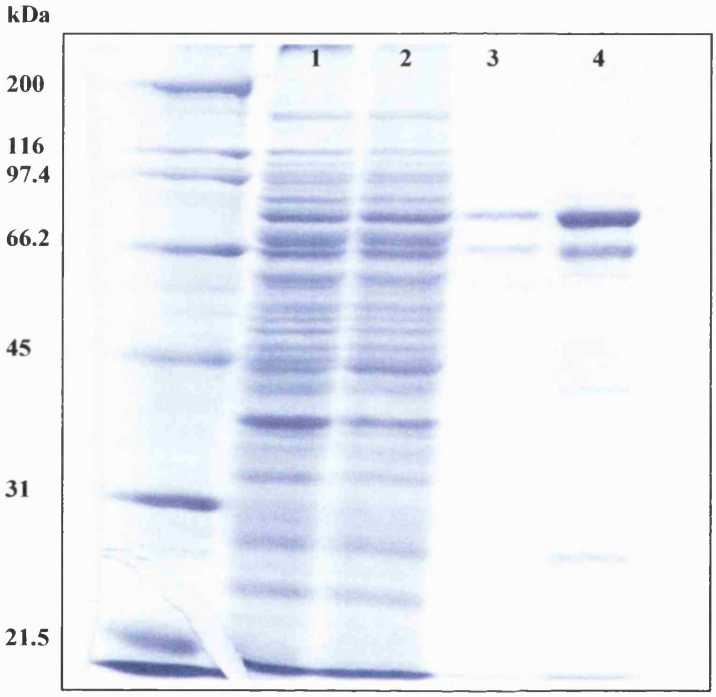


Fig. 6.8: SDS-PAGE analysis of the pullulanase purification.
Lane 1 and 2:Cell extract of BL21 (DE3)/pET28a/Pullulanase;
Lane 3: Sample purified from affinity chromatography;
Lane 4: Sample purified from gel filtration.

6.7.3 Third Stage Purification of Recombinant Pullulanase with Anion Exchange Chromatography

Anion exchange chromatography with Mono Q (H 5/5) column was successful in purifying the recombinant pullulanase. Unbound fractions from the column were found to contain activity when assayed. Pullulanase activity was detected in fractions that were eluted between 0.2-0.22M NaCl. Two fractions from unbound fractions were pooled while three fractions from the eluted fractions were pooled. These fractions are of the highest activity when assayed and were then visualised with SDS PAGE (Fig. 6.9).

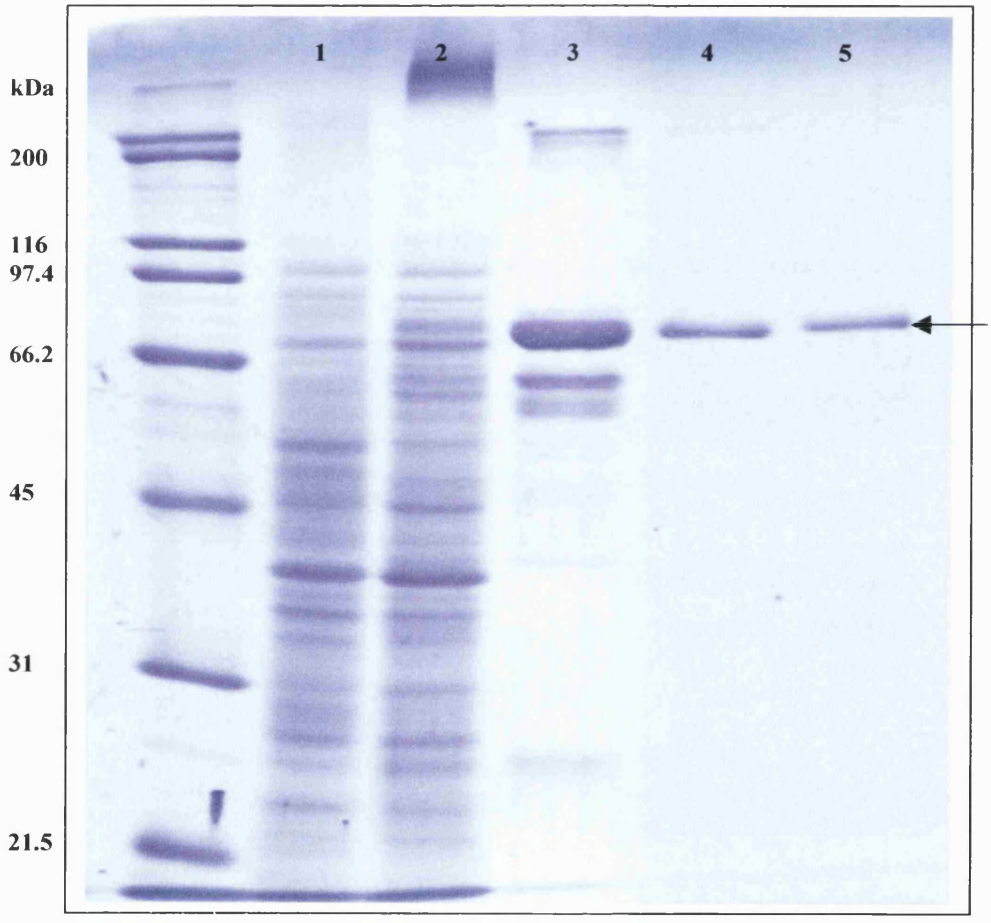


Fig.6.9: SDS-PAGE analysis of anion exchange chromatography with Mono Q (H 5/5) column. Lane 1: Total Protein of BL21 (DE3)/pET28a as negative control; Lane 2: Total Protein of BL21 (DE3)/pET28a/Pullulanase; Lane 3: Samples purified after Affinity Chromatography; Lane 4: Unbound fractions from Anion Exchange Chromatography; Lane 5: Eluted fractions from Anion Exchange Chromatography. The arrow indicates the recombinant pullulanase.

After establishing the methods for purifying recombinant pullulanase, subsequent protein purification was carried out firstly with affinity chromatography and then anion exchange chromatography.

For the affinity purification, several steps were carried out to prevent overloading of the column and any pullulanase that did not bind to the column was subsequently reloaded onto the column. In the following Table 6.6, the affinity chromatography step was carried out four times for the 3ml cell extract. The eluted recombinant pullulanase, from each affinity purification were pooled together to give the total volume of 8ml. Approximately 40% of enzyme activity was found to be the unbound fractions. 4ml of the unbound fractions were reloaded to the affinity column and recombinant pullulanase can still be eluted with 0.2% α-cyclodextrin. Unfortunately, pullulanase can still be detected in the unbound fractions. The % recovery of each run of affinity purification is approximately 12%. The affinity purification approach is only viable for small-scale purification of the recombinant pullulanase. Another approach of purification is needed for large-scale production.

Table 6.6: Purification table of recombinant pullulanase.

	Cell Extract	Affinity Chromatography	Anion Exchange Chromatography
Total volume (ml)	3	8	3
Total Protein (mg)	72.81	0.12	0.017
Enzyme Activity (U/ml)	4.6	0.84	0.66
Total Activity (U)	13.8	6.72	1.98
Specific Activity (U/mg)	0.19	56	116
% Recovery	100	49	14

6.7.4 Determination of the Mr of the Recombinant Pullulanase

Distance travelled by the native pullulanase on the 10% SDS page was found to be 16mm (Fog.6.10). The Mr of the native pullulanase was estimated to be 84 (±8) kDa by solving the quadratic equation produced from the polynomial regression curve in Fig. 6.11. From the gene sequence, the theoretical Mr of the mature pullulanase was calculated to be 94 kDa.

Distance travelled by the recombinant pullulanase on the 10% SDS page was found to be 19.5mm (Fog. 6.12). The Mr of the recombinant pullulanase was estimated to be 77 (±8) kDa by solving the quadratic equation produced from the polynomial regression curve in Fig. 6.13.

As mentioned earlier in Chapter 5, an internal Shine-Dalgarno sequence and an ATG or TTG start sites were found within the pullulanase gene sequence. If protein translation had begun at these sites, the theoretical Mr of the truncated pullulanase would be of 78 kDa and 76 kDa respectively.

Due to the differences produced between the data obtained by Mr estimation from SDS PAGE and theoretical calculation of both native and recombinant pullulanase, it is not possible to conclude whether the recombinant protein or the native protein is the full-length protein or the truncated version. But, by taking into consideration the error limits of SDS PAGE, the native and recombinant pullulanases can be concluded to be similar in size. If the native protein is generated by signal peptide cleavage, it is expected to be present as the mature 94 kDa protein. The Mr of the native protein was determined to be 84 kDa by SDS-PAGE. This indicates that *Thermopallium* pullulanase shows anomalous migration in the system.

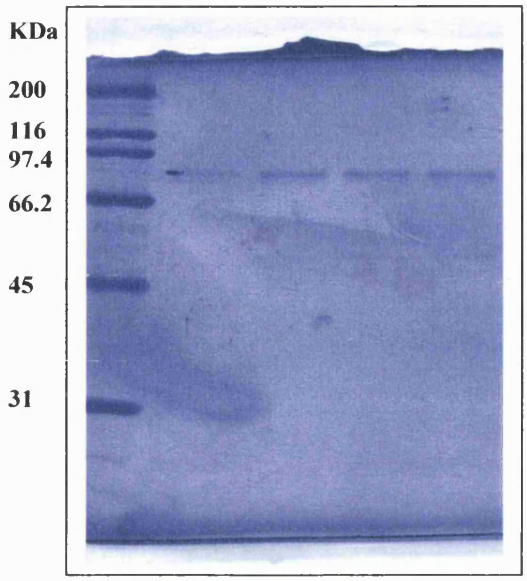


Fig. 6.10: 10% SDS PAGE showing migration of native pullulanase of *T.natronophilum*. All four lanes are of the same enzyme.

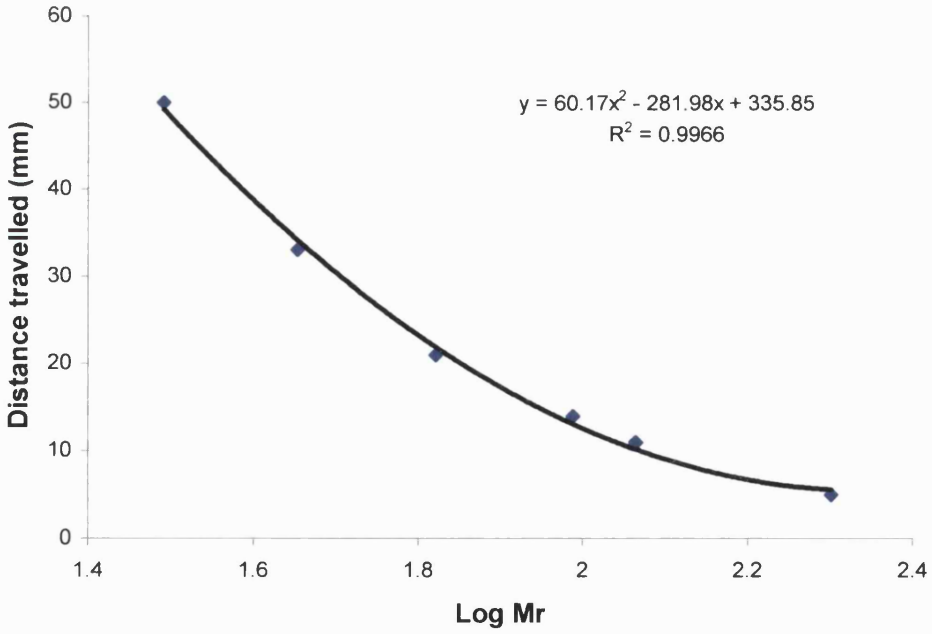


Fig.6.11: Standard curve to estimate the Mr of native pullulanase by SDS-PAGE.

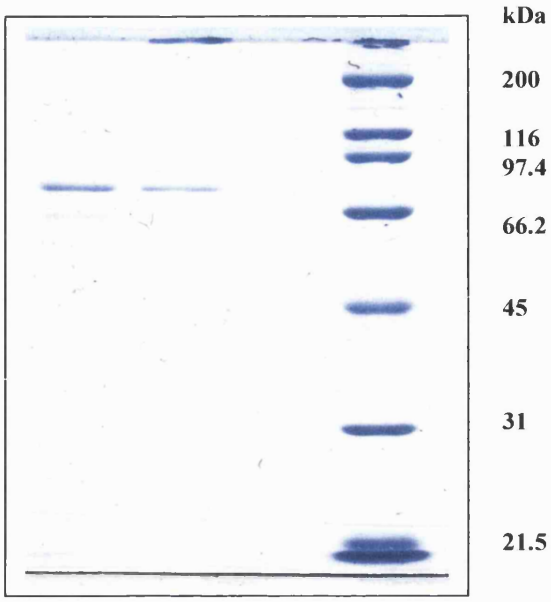


Fig. 6.12: 10% SDS PAGE showing the migration of recombinant pullulanase of *T.natronophilum*.

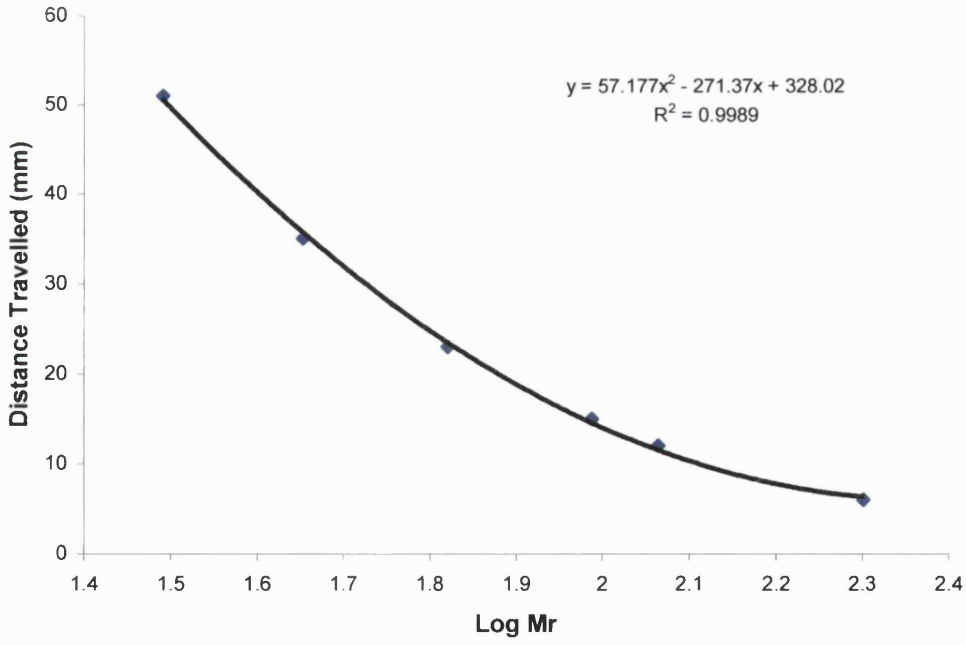


Fig.6.13: Standard curve to estimate the Mr of recombinant pullulanase by SDS-PAGE.

Chapter 7: Characterisation of the Recombinant Pullulanase

7.1 Introduction

After the successful purification of the recombinant pullulanase, characterisation of the recombinant enzyme was carried out. Parameters investigated are temperature optimum, pH optimum, thermal stability, and substrate specificity. Purified native pullulanase was also used in experiments carried out in parallel with the recombinant pullulanase. Comparison of the native and recombinant pullulanase can then be carried out efficiently.

7.2 Methods

7.2.1 Determination of the Temperature Optima for Native and Recombinant Pullulanase

Hydrolytic activity was assayed using the standard assay described in Chapter 2, section 2.2.3. Enzymes were assayed at 50°C, 55°C, 60°C, 65°C, 70°C, 75°C, 80°C, 85°C, 90°C, 95°C and 100°C over a time course of 20 min at each temperature. Assays were carried out in 1.5ml tubes and heated by a micro tube heating block (Grant Instruments, Cambridge, UK).

7.2.2 Determination of the pH Optima for Native and Recombinant Pullulanase

Hydrolytic activity was assayed using the standard assay described in Chapter 2, section 2.2.3. Initial experiments to determine that pH optimum for the native and recombinant pullulanase were carried out with a series of 50mM buffers supplemented with 6.7mM NaCl and 1% pullulan. These buffers are MES, MOPS, HEPES, glycine, tris and borate. The pH shifts caused by the various temperatures

were also compensated for each buffer. A second attempt to investigate the pH optimum was carried out with 50mM sodium phosphate, 6.7mM NaCl with values of 6.0, 6.5,7.0,7.5 and 8.0.at the standard assay temperature. Each buffer was supplied with 1% pullulan as substrate. Assays were carried out in 1.5ml tubes and heated with micro tube heating block (Grant Instruments, Cambridge, UK).

7.2.3 Thermal Inactivation of Native and Recombinant Pullulanase

Aliquots of recombinant and native pullulanase in 0.25ml thin walled PCR tubes (ABgene, Surrey, UK) were heated at temperatures of 80°C, 85°C, 87°C and 90°C in a Mastercyler® (Eppendorf AG, Hamburg, Germany). 30µl aliquots were removed at time points 0, 10, 20, 30 and 60 min and cooled rapidly to 4°C. Hydrolytic activity was then assayed using the standard assay described in Chapter 2, section 2.2.3. Triplicate samples for each time point were taken for each temperature.

7.2.4 Substrate Specificity of Recombinant Pullulanase

Hydrolytic activity was assayed using the standard assay described in Chapter 2, section 2.2.3. 1% amylopectin, 1% amylose and 1% starch were used to replace 1% pullulan in the standard assay.

7.3 Results

7.3.1 Temperature Optima for Native and Recombinant Pullulanase

2.6U of recombinant pullulanase and 0.014U of native pullulanase were used for this experiment. From the results obtained, the optimum temperature for the recombinant pullulanase is 80°C (Fig. 7.1) and the native pullulanase is around 85°C (Fig. 7.2). The experiments were designed with 5°C increment in temperature starting from 50°C to 100°C for recombinant pullulanase and a range of 70°C to 95°C for native pullulanase.

Due to the this 5°C increment, it is therefore possible that 80°C might not be the true temperature optimum for recombinant pullulanase as in the case of 85°C for native pullulanase. However, the temperature optimum for recombinant pullulanase is in the range of 77°C to 83°C and similarly for the case of native pullulanase in the range of 82°C and 87°C.

One possible explanation for the difference in temperature optima between the two enzymes is that the recombinant pullulanase is not as stable as the native pullulanase at high temperature. Further experiments that investigate the enzyme activity between the range of 75°C to 85°C for the recombinant and native pullulanase have to be carried out before more precise conclusions can be drawn.

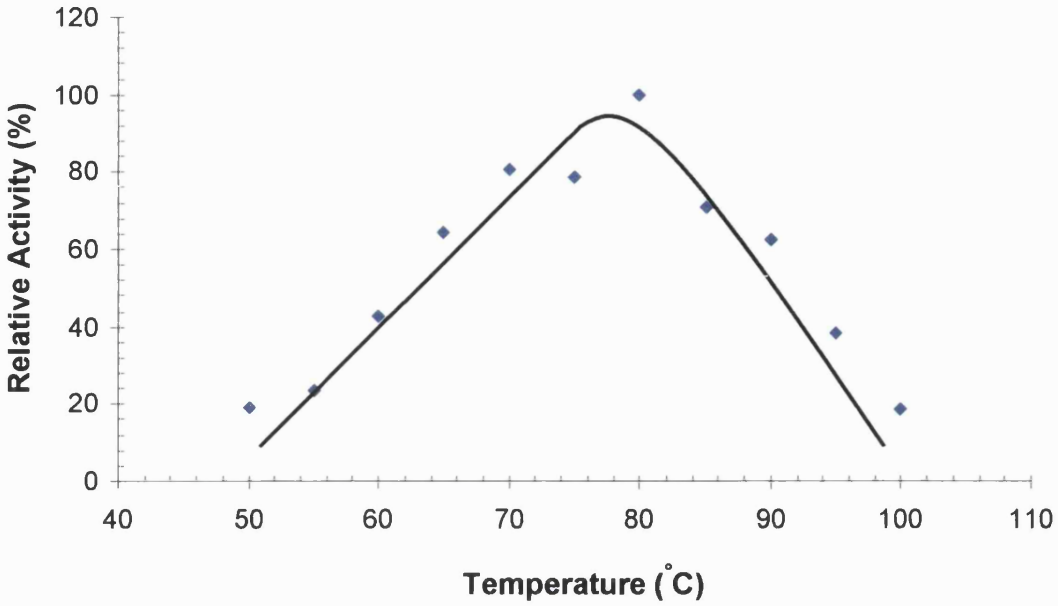


Fig.7.1: Temperature optimum for recombinant pullulanase.

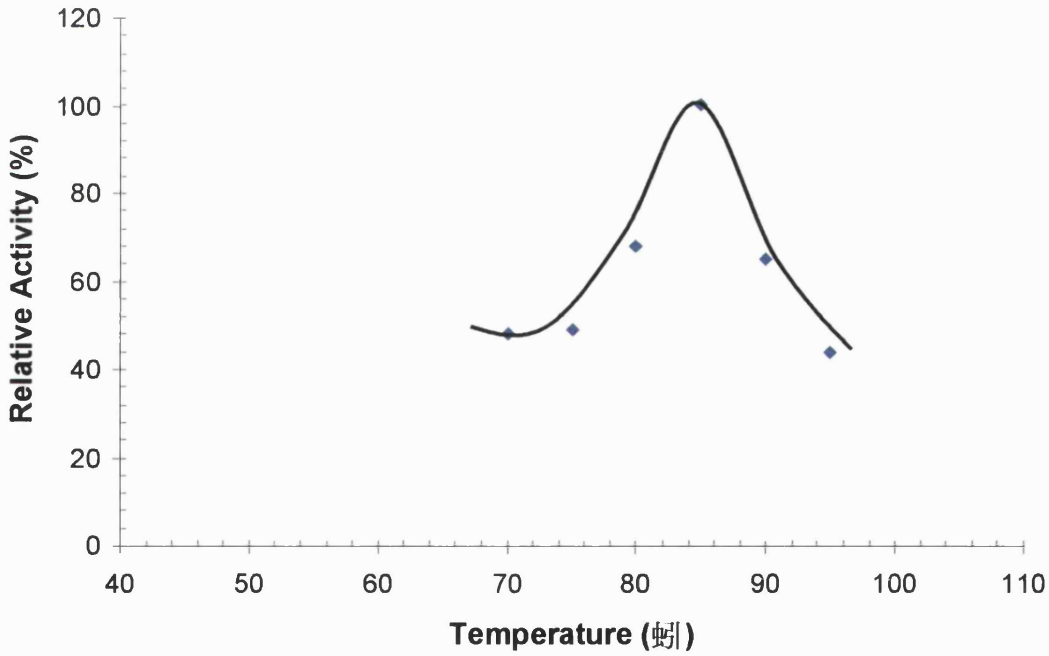


Fig.7.2: Temperature optimum for native pullulanase

7.3.2 pH Optima for Native and Recombinant Pullulanase

1.8U of recombinant pullulanase and 0.014U of native pullulanase were used for this experiment. From the results obtained, the optimum pH for the recombinant and native pullulanase is pH 7 (Fig. 7.3 and Fig. 7.4)

In preliminary experiments, the pH optima of both pullulanase were investigated using a series of buffers with different pH values. Buffers used were MES pH6, pH6.8, MOPS pH6.7, pH7.3, pH8.1, HEPES pH8, pH8.6, Tris pH9, pH9.1, Borate pH9.3, pH10, Glycine pH10.2, pH10.9, pH11.3. With this series of different buffer used at 50mM concentration, the optimum pH of the recombinant pullulanase is found to be pH 6.7 (Fig. 7.5) and the optimum pH of the native pullulanase is found to be at pH6.8 (Fig. 7.6). The shift in pH corresponding to the increase in temperature (assay condition at 80°C) of each of these chemicals has been taken into consideration.

Table 7.1: Buffers used with their pKa values, useful pH range and pH shift with an increase in 1°C in temperature (Eisenthal and Danson, Enzyme Assays, 2nd Ed.).

Buffers	pKa (25°C)	Useful pH Range	δpH/δT
MES	6.1	5.1-7.1	-0.011/°C
MOPS	7.2	6.2-8.2	-0/015/°C
HEPES	7.5	6.5-8.5	-0.014/°C
Glycine	9.6	8.6-10.6	-0.025/°C
Tris	8.1	7.1-9.1	-0.031/°C
Borate	9.2	8.2-10.2	-0.008/°C
NaH_2PO_4	7.2	6.2-8.2	-0.0082/°C

By comparing the sets of data obtained from two different sets of buffers, the optimum pH for both the native and recombinant pullulanase is round pH 7. Thus, a conclusion can safely be drawn that this is indeed the true pH optimum of the enzyme and that the recombinant enzyme has similar properties to the native pullulanase.

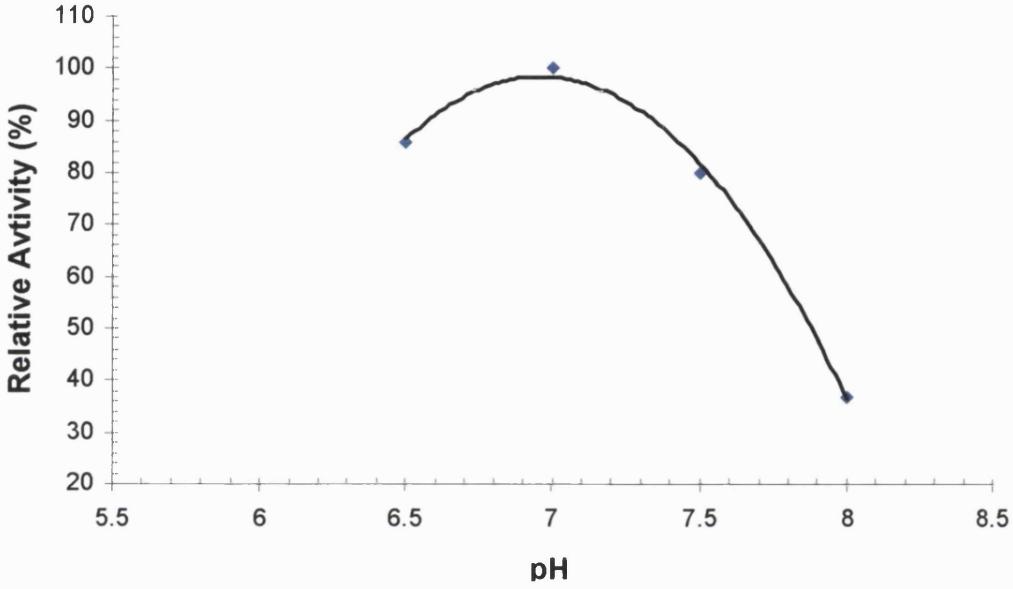


Fig.7.3: pH optimum for recombinant pullulanase using buffer NaH_2PO_4.

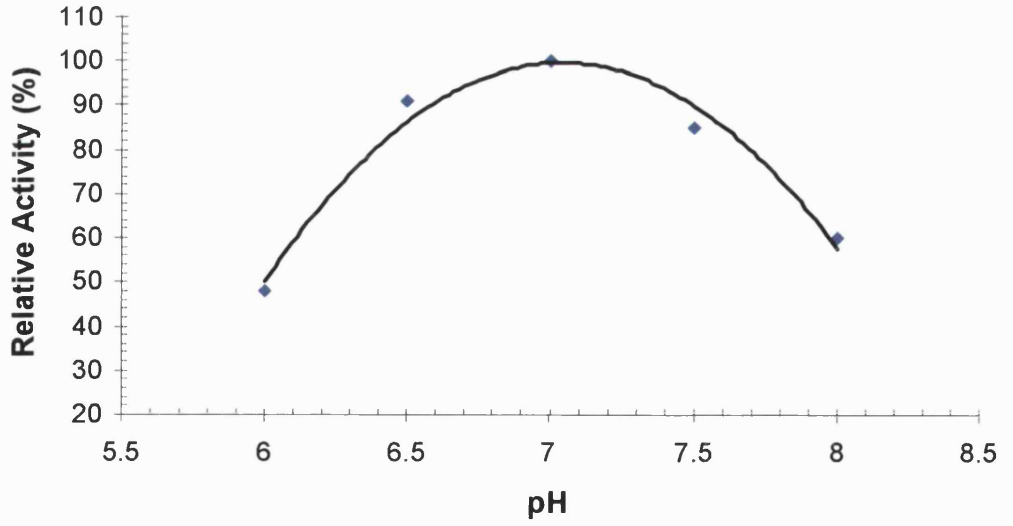


Fig.7.4: pH optimum for native pullulanase using buffer NaH_2PO_4.

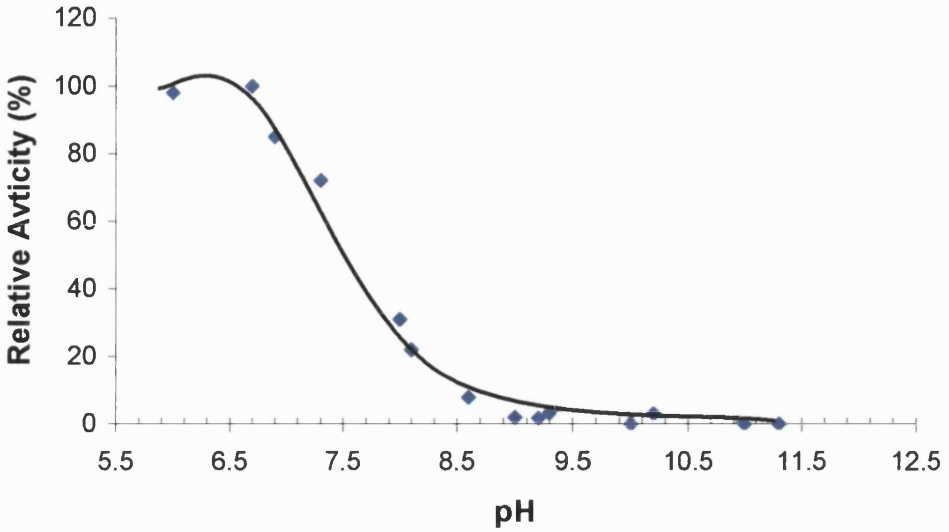


Fig. 7.4: pH optima of recombinant pullulanase determined using buffers MES, MOPS, HEPES, Glycine. Tris and Borate.

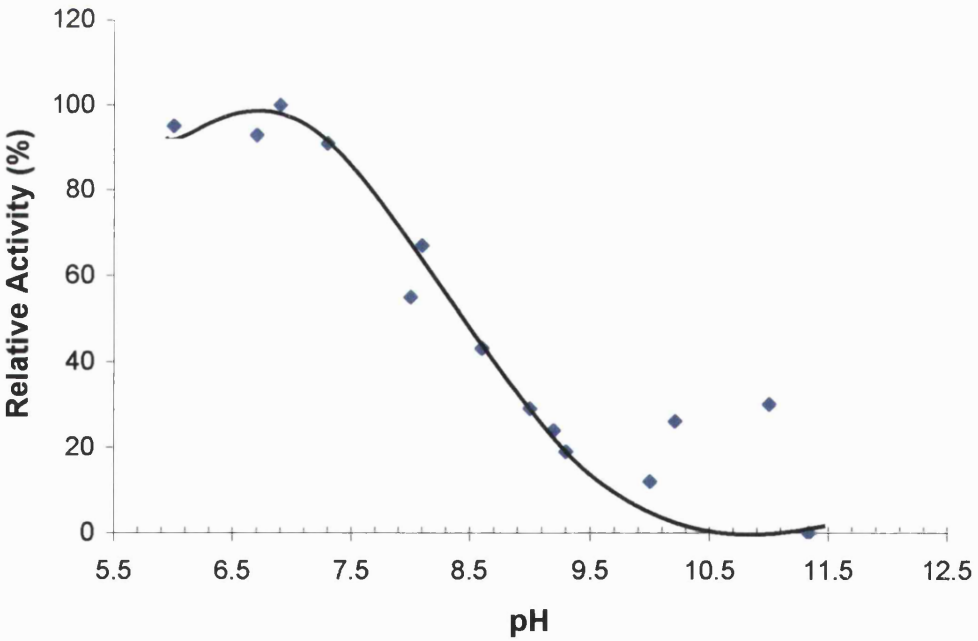


Fig. 7.5: pH optima of native pullulanase determined using buffers MES, MOPS, HEPES, Glycine. Tris and Borate.

7.3.3 Thermal Inactivation Studies on Native and Recombinant Pullulanase

2.6U of recombinant pullulanase and 0.014U of native pullulanase were used for this experiment.

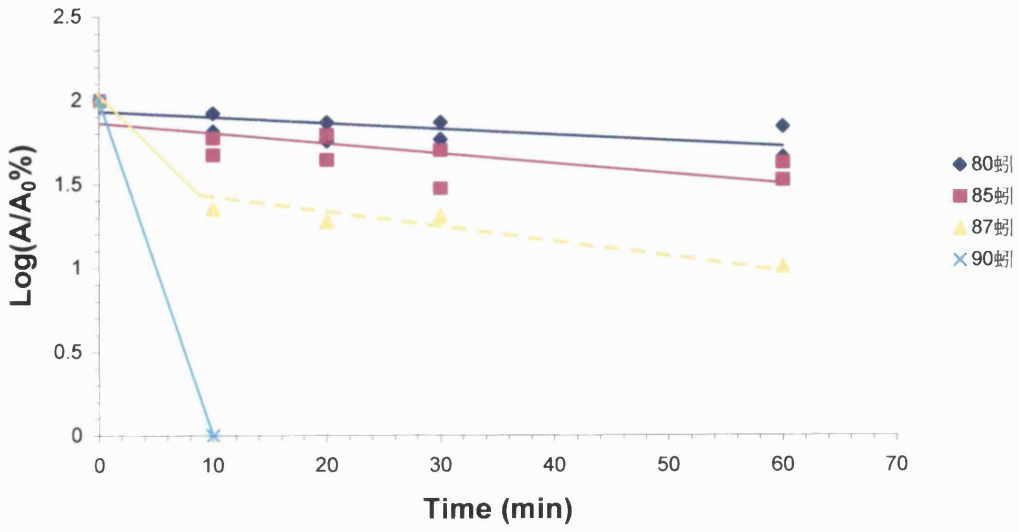


Fig. 7.6: Thermal inactivation studies on recombinant pullulanase

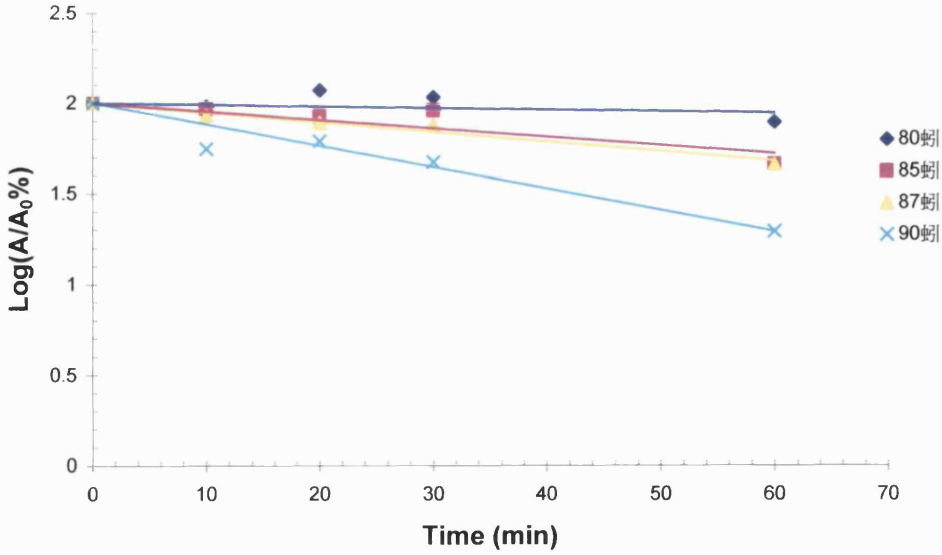


Fig. 7.7: Thermal inactivation studies on native pullulanase

By comparing the graphs in Fig. 7.6 and Fig. 7.7, which depict the thermal stability of the two enzymes, it is found that the native enzyme is more thermostable than the recombinant pullulanase. This is clearly seen at 90°C, where the native enzyme is slowly inactivated but the recombinant pullulanase is already completely inactivated after 10 min incubation. Following this preliminary observation, the rate constants for the inactivation of each enzyme at each temperature was calculated based on the formula:

Equation 1

$$\ln A = \ln A_0 - kt$$

$$\therefore \ln(A/A_0) = -kt$$

$$\therefore 2.303 \log(A/A_0\%) = -kt$$

whereby,

t= time
k=rate constant
A=activity at time t
A_0= activity when t=0 (intercept of y-axis)

The rate constants for inactivation of both pullulanase at each temperature are shown in Table 7.2. From here, the half-life of the enzyme at each temperature (Table 7.2) is obtained by substituting rate constants obtained into equation 2 (derived from equation 1).

Equation 2

$$\text{when } (A/A_0\%) = 0.5$$

$$\text{then, } 2.303 \log 0.5 = -kt_{1/2}$$

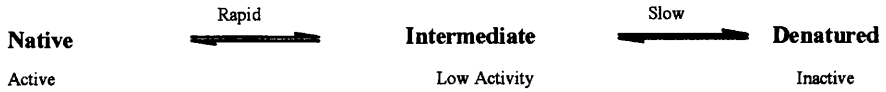
$$\therefore t_{1/2} = 0.6932/k$$

Table 7.2: Thermal inactivation of the recombinant and native pullulanase.

Temperature	Rate Constant (min^{-1})		Half-life (min)	
	Recombinant Pullulanase	Native Pullulanase	Recombinant Pullulanase	Native Pullulanase
80°C	0.0081	0.0041	86	169
85°C	0.0140	0.0129	50	54
87°C	-	0.0127	-	55
90°C	0.4606	0.0251	1.5	28

From the half-life of both recombinant and native pullulanase at each temperature (Table 7.2), the conclusion can be drawn that the native pullulanase is more thermostable than the recombinant pullulanase at each temperature investigated. The recombinant pullulanase showed a slower rate of loss of activity for temperatures at 80°C and 85°C. The recombinant pullulanase showed a rapid loss of activity for the first 10 min for thermal inactivation studies at 87°C but seems to stabilise and show a slower rate of activity loss (yellow dotted line) after the first 10 min (Fig. 7.6). This pattern of activity loss is observed in thermal inactivation studies of porcine heart citrate synthase (McEvily and Harrison, 1986), whereby the initial loss of activity was caused by dissociation of the dimers. *T.natronophilum* pullulanase is a monomeric protein with at least thee domains or more, which might explain the initial loss of activity due to domain shifts as the temperature increases. The native pullulanase is not showing such a pattern of activity loss. The recombinant pullulanase is genetically modified whereby the predicted signal peptide of the protein was removed before cloning. There is a possibility that the predicted cleavage site of the signal peptide is not true for the native pullulanase. The removal of some of the amino acids might affect the thermal stability of the recombinant pullulanase.

The recombinant protein might unfold rapidly to an intermediate, which stabilises out but showed reduced enzyme activity. This can be explained by the following equation,



It is also possible that the presence or absence of a signal peptide affects the folding of the enzyme. In this case the native precursor would fold differently from the recombinant mature protein.

Thermal inactivation studies were also carried out on unpurified crude samples of the recombinant enzyme at temperatures 75°C and 85°C. This work was carried out by Elizebeth Hodder, a final year undergraduate project student. The half-life of pullulanase in unpurified crude extracts at 85°C is 54 min, which is in good agreement with the half-life obtained for purified recombinant pullulanase. This shows that the purification steps did not alter or affect the characteristics of the recombinant pullulanase.

7.3.4 Substrate Specificity of Recombinant Pullulanase

Recombinant pullulanase was assayed with different substrates to determine its substrate specificity. The results obtained are shown in Fig. 7.8. As expected, the recombinant pullulanase showed the highest activity with pullulan (100%), followed by amylopectin (7.5%) and starch (1%). The results obtained further established the characteristics of this recombinant pullulanase to be a true type I pullulanase. The results obtained agree with the results obtained for native pullulanase, which hydrolyses pullulan, and showed a slight enzyme activity when incubated with starch and amylopectin. The substrate specificity determination of native pullulanase was carried out by Thompson (1998).

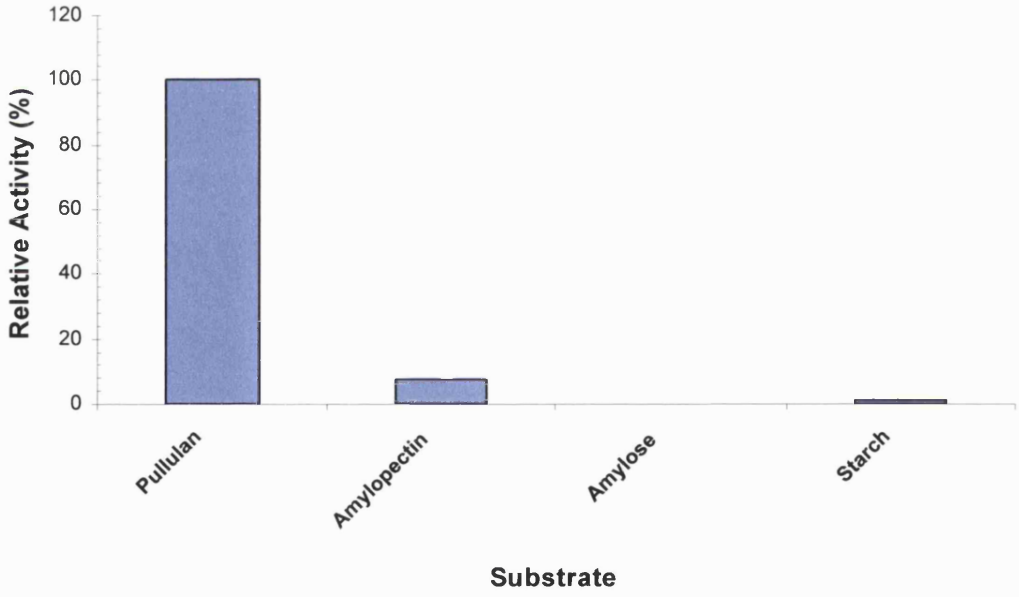


Fig. 7.8: Substrate specificity of recombinant pullulanase. 1% of each substrate was used in the standard assays.

General Conclusions and Proposals for Future Work

In conclusion, the gene sequence of the pullulanase type I from *Thermopallium natronophilum* have been successfully determined. The cloning and expression of the recombinant pullulanase in an *E. coli* host were also successful. As for the purification of the recombinant protein, affinity chromatography is an effective step although it is not ideal if large amounts of protein are needed for industrial use.

The purified recombinant pullulanase showed similar substrate specificity to the native enzyme and had a similar pH optimum. However, the recombinant enzyme showed reduced thermal stability, and the basis for this has still to be investigated. The size of the native and recombinant pullulanases were similar, and it is most likely that in each case this represents the mature protein sequence, which lacks the N-terminal signal peptide identified in the pullulanase gene. However, it remains possible that in the recombinant a truncated protein is expressed from two internal Shine-Dalgarno sequences and TTG/ATG start sites. N terminal sequencing of the expressed recombinant protein could be carried out in order to resolve this anomaly. Considering the fact that there are no three-dimensional structures of a thermostable type I pullulanases currently available, a venture to obtain crystals and solve the structure of this enzyme should be attempted once the anomaly mentioned above has been resolved.

Mutational studies can be carried out on the highly conserved peptide sequence (YNWGYDP) found in several type I pullulanases to investigate the role of each amino acid in substrate binding and hydrolysis of α-1,6 glycosidic linkages. Furthermore, mutational studies should also be carried out on amino acids predicted in Chapter 5 to be involved in substrate catalysis.

Both native and recombinant pullulanase have pH optima of 7, which is lower than the optimum required for current applications in the detergent industry. To improve the potential of this enzyme for application in the detergent industry, directed

evolution studies could be carried out to identify a mutant enzyme that is highly active at high pH.

The detergent industry is moving towards cleaning at lower temperature. In Europe, societies are encouraged to carry out cleaning at temperatures between 40°C and 60°C in order to save energy. In Japan, household cleaning is already carried out at room temperature. Considering the shift of interest in washing temperature, directed evolution can also be carried out to look for mutants with high enzyme activity at lower temperature and high pH. Only till all these are established, can this enzyme be assessed for its suitability in the cleaning industry.

References

Albertson, G. D., McHale, R. H., Gibbs, M. D. and Bergquist, P. L. (1997) Cloning and sequence of a type I pullulanase from an extremely thermophilic anaerobic bacterieum *Caldicellulosiruptor saccharolyticus*. *Biochim. Biophys. Acta* **1354**: 35-39.

Aoki, H., Sakano, Y. and Sakano, Y. (1997) Molecular cloning and heterologous expression of the isopullulanase gene from Aspergillus niger A.T.C.C. 9642. *Biochem. J.* **323**: 757–764.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. and Sonnhammer, E. L. L. (2002) The Pfam protein families database. *Nucleic Acids Res.* **30**(1): 276-280.

Bernfeld, P. (1955) Amylases, α and β. *Methods Enzymol.* **1**: 149-155.

Bertoldo, C. and Antranikian, G. (2001). Amylolytic enzymes from hyperthermophiles. *Methods in Enzymology: Hyperthermophilic Enzyme PTA.* **330:** 269-289.

Bertoldo, C. and Antranikian, G. (2002) Starch-hydrolyzing enzymes from thermophilic archae and bacteria. *Curr. Opin. Chem. Biol.* **6**: 151-160.

Bertoldo, C., Duffner, F., Jorgensen, P. L. and Antranikian, G. (1999) Pullulanase type I from *Fervidobacterium pennavorans* Ven5: Cloning, sequencing, and expression of the gene and biochemical characterization of the recombinant enzyme. *Appl. Environ. Microbiol.* **65**(5): 2084–2091.

Bibel, M., Brettl, C., Gosslar, U., Kriegshäuser, G. and Liebl, W. (1998) Isolation and analysis of genes for amylolytic enzymes of the hyperthermophilic bacterium *Thermotoga maritima*. *FEMS Microb. Letts.* **158**: 9-15.

Brzozowski A.M. and Davies, G. J. (1997) Structure of the Asperfilus oryzae a-amylase complexed with the inhiitor acarbose 2.0A resolution. *Biochemistry*(36): 10837-10845.

Brzozowski, A. M., Lawson, D. M., Turkenburg, J. P., Bisgaard-Frantzen, H., Svendsen, A., Borchert, T. V., Dauter, Z., Wilson, K. S. and Davies, G. J. (2000) Structural analysis of a chimeric bacterial α-amylase. High-resolution analysis of native and ligand complexes. *Biochemistry* **39**: 9099-9107.

Buléon, A., Colonna, P., Planchot, V. and Ball, S. (1998) Starch granules: structure and biosymthesis. *Intern. J. Biol. Macromol.* **23**: 85-112.

Burgess, A. W., Ponnuswamy, P. K. and Sheraga, H. A. (1974) Analysis of conformations of amino acid residues and prediction of backbone topography in proteins. *Israel J. Chem.* **12**: 239-286.

Davies, G. J. and Wilson, K. S. (1999) Trapped in the act of catalysis. *Nat. Struct. Biol.* **6**(5): 406-408.

Davies, G. J., Wilson, K. S. and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* **321**: 557-559.

Dele`age, G. and Roux, B. (1987) An algorithm for secondary structure prediction based on class prediction. *Prot. Engineer.* **1**(4): 289-294.

Duckworth, A. W., Grant, W. D., Jones, B. E. and Van Steenbergen, R. (1996) Phylogenetic diversity of soda lake alkaliphiles. *FEMS Microbial. Ecol.* **19**: 181-191.

Duffner, F., Bertoldo, C., Anderson, J. T., Wagner, K. and Antranikian, G. (2000) A new thermoactive pullulanase from *Desulfurococcus mucosus*: Cloning, sequencing, purification and characterization of the recombinant enzyme after expression in *Bacillus subtilis*. *J. Bacteriol.* **182**(22): 6331-6338.

Eisenthal, R. and Danson, M.J. (2003) *Enzyme Assays*, 2nd Edition, Oxford University Press.

Fersht, A. (1999). *Structure and Mechanism in Protein Science.*, W.H.Freeman and Company.

Frantzen, H. B. and Svendsen, A. (2002). Starch debranching enzymes. *Patent: JP 2002519054-A 2, Novozymes AS.*

Gantelet, H. and Duchiron, F. (1999) A new pullulanase from a hyperthermophilic archaeon for starch hydrolysis. *Biotech. Letts.* **21**: 71-75.

Gantelet, H., Ladrat, C., Godfroy, A., Barbier, G. and Duchiron, F. (1998) Characteristics of pullulanases from extremely thermophilic archaea isolated from deep-sea hydrothermal vents. *Biotech. Letts.* **20**(9): 819-823.

Garnier, J., Gibrat, J. F. and Robson, B. (1996). GOR method for predicting protein secondary structure from amino acid sequence. *Meth. Enzym.* Doolittle, R. F. **266:** 97-120.

Garnier, J., Levin, J. M., Gibrat, J. F. and Biou, V. (1990) Secondary structure prediction and protein design. *Biochem. Soc. Symp.* **57**: 11-24.

Garnier, J., Osguthorpe, D. J. and Robson, B. (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of proteins. *J. Mol. Biol.* **120**: 97-120.

Hatada, Y., Igarashi, K., Ozaki, K., Ara, K., Hitomi, J., Kobayashi, T., Kawai, S., Watabe, T. and Ito, S. (1996) Amino acid sequence and molecular structure of an alkaline amylopullulanase from *Bacillus* that hydrolyzes α-1,4 and α-1,6 linkages in polysaccharides at differenct active sites. *J. Biol. Chem.* **271**(39): 24075-24083.

Hatada, Y., Saito, K., Hagihara, H., Ozaki, K. and Ito, S. (2001) Nucleotide and deduced amino acid sequences of an alkaline pullulanase from the alkaliphilic bacterium *Bacillis sp.* KSM-1876. *Biochim. Biophys. Acta* **1545**: 367-371.

Holley, L. H. and Karplus, M. (1989) Protein secondary structure prediction with a neural network. *Proc. Natl. Acad. Sci. USA* **86**: 152-156.

Hondoh, H., Kuriki, T. and Matsuura, Y. (2003) Three-dimensional structure and substrate binding of *Bacillus stearothermophilus* neopullulanase. *J. Mol. Biol.* **326**: 177-188.

Horinouchi, S., Fukusumi, S., Ohshima, T. and Beppu, T. (1988) Cloning and expression in *Escherichia coli* of two additional amylase genes of a strictly anaerobic thermophile, *Dictyoglomus thermophilum*, and their nucleotide sequences with extremely low guanine-plus-cytosine contents. *Eur. J. Biochem.* **176**(2): 243-53.

Huber, R., Langworthy, T. A., König, H., Thomm, M., Woese, C. R., Sleytr, U. B. and Stetter, K. O. (1986) *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch. Microbiol.* **144**: 324-333.

Ito, S., Kobayashi, T., Ara, K., Ozaki, K., Kawai, S. and Hatada, Y. (1998) Alkaline detergent enzymes from alkaliphiles: enzymatic properties, genetics and structures. *Extremophiles* **2**: 185-190.

Janeck, S. (1992) New conserved amino acid region of α-amylases in the third loop of their $(\beta/\alpha)_8$-barrel domains. *Biochem. J.* **288**: 1069-1075.

Janse, B. J. H., Steyn, A. J. C. and Pretorius, I. S. (1993) Regional sequence homologies in starch-degrading enzymes. *Curr. Genet.* **24**: 400-407.

Jespersen, H. M., MacGregor, E. A., Henrissat, B., Sierks, M. R. and Svensson, B. (1993) Starch- and glycogen-debranching and branching enzymes: prediction of structural features of the catalytic $(\beta/\alpha)_8$-barrel domain and evolutionary relationship to other amylolytic enzymes. *J. Prot. Chem.* **12**(6): 791-805.

Jespersen, H. M., MacGregor, E. A., Sierks, M. R. and Svensson, B. (1991) Comparison of the domain-level organization of starch hydrolases and related enzymes. *Biochem. J.* **280**: 51-55.

Jones, B. E., Grant, W. D., Duckworth, A. W. and Owenson, G. G. (1998) Microbial diversity of soda lakes. *Extremophiles* **2**: 191-200.

Kainuma, K., Kobayashi, S. and Harada, T. (1978) Action of Pseudomonas isoamylase on various branched oligo- and polysaccharides. *Carbohydr. Res.* **61**: 345-357.

Kamitori, S., Kondo, S., Okuyama, K., Yokota, T., Shimura, Y., Tonozuka, T. and Sakano, Y. (1999) Crystal structure of *Thermoactinomyces vulgaris* R-47 α-amylase II (TVAII) hydrolyzing cyclodextrins and pullulanase at 2.6Å resolution. *J. Mol. Biol.* **287**: 907-921.

Kashiwabara, S., Ogawa, S., Miyoshi, N., Oda, M. and Suzuki, Y. (1999) Three domains comprised in thermostable molecular weight 54,000 pullulanase type I from *Bacillus flavocaldarius* KP1228. *Biosci. Biotech. Biochem.* **63**(10): 1736-1748.

Katsuragi, N., Takizawa, N. and Murooka, Y. (1987) Entire nucleotide sequence of the pullulanase gene of *Klebsiella aerogenes* W70. *J. Bacteriol.* **169**: 2301-2306.

Khajeh, K., Ranjbar, B., Naderi-Manesh, H., Habibi, A. E. and Nemat-Gorgani, M. (2001) Chemical modification of bacteriala α-mylases: changes in tertiary structures and the effect of additional calcium. *Biochim. Biophys. Acta.* **1548**: 229-237.

Kim, C., Choi, H. and Lee, D. (1993) Pullulanases of alkaline and broad pH range from a newly isolated alkalophilic *Bacillus sp.* S-1 and a *Micrococcus sp.* Y-1. *J. Indust. Microbiol.* **12**: 48-57.

Kim, C., Choi, H. and Lee, D. (1993) Purification and biochemical properties of an alkaline pullulanase from alkalophilic *Bacillus sp.* S-1. *Biosci. Biotech. Biochem.* **57**(10): 1632-1637.

Kim, C., Nashiru, O. and Ko, J. H. (1996) Purification and biochemical characterization of pullulanase type I from *Thermus caldophilus* GK-24. *FEMS Microb. Letts.* **138**: 147-152.

Kim, T., Park, W., Choi, S., Yoon, Y., Park, K. and Kim, J. (2000) Molecular cloning and characterization of a thermostable pullulanase from a *Thermus* Strain IM6501. *Food Sci. Biotechnol.* **9**(3): 188-194.

King, R. D. and Sternberg, M. J. E. (1990) Machine learning approach for the prediction of protein secondary structure. *J. Mol. Biol.* **216**: 441-457.

King, R. D. and Sternberg, M. J. E. (1996) Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Science* **5**: 2298-2310.

Klein, C., Hollender, J., Bender, H. and Schulz, G. E. (1992) Catalytic center of cyclodextrin glycosyltransferase derived from X-ray structure analysis combined with site-directed mutagenesis. *Biochemistry* **31**: 9740-8746.

Kriegshäuser, G. and Liebl, W. (2000) Pullulanase from the hyperthermophilic bacterium *Thermotoga maritima*: purification by β-cycledextrin affinity chromatography. *J. Chroma. B* **737**: 245-251.

Krik, O., Borchert, T. V. and Fuglsang, C. C. (2002) Industrial enzyme applications. *Curr. Opin. Biotechnol.* **13**: 345-351.

Kuriki, T., Takata, H., Okada, S. and Imanaka, T. (1991) Analysis of the active centre of *Bacillus stearothermophilus* neopullulanase. *Journal of Bacteriology* **173**(19): 6147-6152.

Lebbink, J. H. G., Bertoldo, C., Tibbelin, G., Anderson, J. T., Antranikian, G. and Ladenstein, R. (2000) Crystallization and preliminary X-ray crystallographic studies of the thermoactive pullulanase type I, hydrolyzing α-1,6 glycosidic linkages from *Fervidobacterium pennivorans* Ven5. *Acta. Cryst. D* **56**: 1470-1472.

Lee, M., Lee, Y. and Kim, C. (1997) Intracellular and extracellular forms of alkaline pullulanase from an alkaliphilic *Baillus sp.* S-1. *Arch. Biochem. Biophys.* **337**(2): 308-316.

Leemhuis, H., H.J., R., Dijkstra, B. W. and Dijkhuizen, L. (2003) The fully conserved Asp residue in conserved sequence region I of the α-amylase family is crucial for the catalytic site architecture and activity. *FEBS Letters* **541**: 47-51.

Liebl, W., Stemplinger, I. and Ruile, P. (1997) Properties and gene structure of the *Thermotoga maritima* α-amylase AmyA, a putative lipoprotein of a hyperthermophilic bacterium. *J. Bacteriol.* **179**(3): 941-948.

Lu, X. Y., Jiang, R. Z. and Wang, G. F. (1991) Molecular cloning of alpha-amylase gene from *Bacillus megaterium* and its expression in *Bacillus subtilis*. *Yi Chuan Xue Bao* **18**(2): 185-92.

MacGregor, E. A., Janeck, S. and Svensson, B. (2001) Relationship of sequence and structure to specificity in the α-amylase family of enzymes. *Biochim. Biophys. Acta* **1546**: 1-20.

Machius, M., Wiegand, G. and Huber, R. (1995) Crystal structure of calcium-depleted *Bacillus licheniformis* α-amylase at 2.2Å resolution. *J. Mol. Biol.* **246**: 545-559.

Matsuura, Y., Kusunoki, M., Harada, W. and Kakudo, M. (1984) Structure and possible catalytic residues of Taka-amylase A. *J. Biochem.* **95**: 697.

McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Körner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W. and Withers, S. G. (1996) The p*Ka* of the general acid/base carboxyl group of a glycosidase cycles during catalysis: A ^{13}C-NMR study of *Bacillus circulans* xylanase. *Biochemistry* **35**: 9958-9966.

Messaoud, E. B., Ammar, Y. B., Mellouli, L. and Bejar, S. (2002) Thermostable pullulanase type I frm new isolated *Bacillus thermoleovorans* US105: cloning, sequencing and expression of the gene in *E.coli*. *Enz. Microbial. Tech.* **31**: 827-832.

Nakamura, A., Haga, K. and Yamane, K. (1993) Three histidine residues in the active center of cyclodextrin glucanotransferase from alkalophilic *Bacillus sp.* 1011: Effects of the replacement on pH dependence and transition-state stabilization. *Biochemistry* **32**: 6624-6631.

Nielsen, H., Engelbrecht, J., Brunak, S. and Heijne, G. V. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Prot. Eng.* **10**.

Nielsen, J. E. and Borchert, T. V. (2000) Protein engineering of bacterial -amylases. *Biochim. Biophys. Acta* **1543**(2): 253-274.

Niven, G. W. and Scurlock, P. G. (1993) A method for the continous purification of proteins by affinity adsorption. *J. Biotech.* **31**: 179-190.

Panasik Jr., N., Brenchley, J. E. and Farber, G. K. (2000) Distributions of structural features contributing to thermostability in mesophilic and thermophilic a/b barrel glycosyl hydrolases. *Biochim. Biophysic. Acta* **1543**: 189-201.

Park, K., Kim, T., Cheong, T., Kim, J., Oh, B. and Svensson, B. (2000) Structure, specificity and function of cyclomaltodextrinase, a multispecific enzyme of the -amylase family. *Biochim. Biophys. Acta* **1478**(2): 165-185.

Pugsley, A. P., Chapon, C. and Schwartz, M. (1986) Extracellular pullulanase of Klensiella pneumoniae is a lipoprotein. *J. Bacteriol.* **166**: 1083-1088.

Rachel, R., Engel, A. M., Huber, R., Stetter, K. O. and Baumeister, W. (1990) A porin-type protein is the main constituent of the cell envelope of the ancestral eubacterium *Thermotoga maritima*. *FEBS Letts.* **262**(1): 64-68.

Raven, N., Ladwa, N., Cossar, D. and Sharp, R. (1992) Continous culture of the hyperthermophilic archaeum *Pyrcoccus furiosus*. *Appl. Microbiol. Biotechnol.* **38**: 263-267.

Rüdiger, A., Jorgensen, P. L. and Antranikian, G. (1995) Isolation and chracterization of a heat-stable pullulanase from the hyperthermophilic archaeon *Pyrococcus woesei* after cloning and expression of its gene in *Escherichia coli*. *Appl. Environ. Microbiol.* **61**(2): 567-575.

Rye, C. S. and Withers, S. G. (2000) Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **4**: 573-580.

Saha, B. C., Mathupala, S. P. and Zeikus, J. G. (1991). Comparison of amylopullulanase to α-amylase and pullulanase. *American Chemical Society Symposium*: 362-371.

Sauer, J., Sigurskjold, B., Christensen, U., Frandsen, T., Mirgorodskaya, E., Harrison, M., Roepstorff, P. and Svensson, B. (2000) Glucoamylase: structure/function relationships, and protein engineering. *Biochim. Biophys. Acta* **1543**(2): 275-293.

Schiraldi, C. and De Rosa, M. (2002) The production of biocatalysts and biomolecules from extremophiles. *Trends. Biotechnol.* **12**: 515-21.

Søgaard, M., Kadziola, A., Haser, R. and Svensson, B. (1993) Site-directed mutagenesis of Histidine 93, Aspartic acid 180, Glutamic acid 205, Histidine 290 and Aspartic acid 291 at the active site and Tryptophan 279 at the raw starch binding site in barley α-amylase 1. *J. Biol. Chem.* **286**(30): 22480-22484.

Sowers, K. R. and Schreier, H. J., Eds. (1995) *Archaea - A Laboratory Manual: Methanogens*. NY, Cold Spring Harbor Press.

Stefanova, M., Schwerdtfeger, R., Antranikian, G. and Scandurra, R. (1999) Heat-stable pullulanase from Bacillus acidopullulyticus: characterization and refolding after guanidinium chloride-induced unfolding. *Extremophiles* **3**(2): 147-152.

Strokopytov B., Penninga D., Rozeboom H.J., Kalk K.H., L., D. and B.W., D. (1995) X-ray structure of cyclodextrin glycosyltransferase complexed with acarbose. Implication for the catalytic mechanism of glycosidases. *Biochemistry* **34**: 2234-2240.

Sugimura K. and Nishihara T. (1988) Purification, characteirazation and primary structure of *E.coli* protease VII with specificity for paired basic resides: Identity of protease VII and OpmT. *J. Bacteriol.* **170** (12): 5625-5632.

Suzuki, Y., Hatagaki, K. and Oda, H. (1991) A hyperthermostable pullulanase produced by an extreme thermophile, Bacillus flavocaldarius KP 1228, and evidence for the proline theory of increasing protein thermostability. *Appl. Microbiol. Biotechnol.* **34**(6): 707-714.

Svensson, B. (1994) Protein engineering in the α-amylase family: Catalytic mechanism, substrate specificity, and stability. *Plant Mol. Biol.* **25**: 141-157.

Svensson, B. and Søgaard, M. (1993) Mutational analysis of glycosylase function. *J. Biotech.* **29**: 1-37.

Takase, K. (1992) Interaction of catalytic-site mutants of *Bacillus subtilis* α-amylase with substrates and acarbose. *Biochim. Biophys. Acta* **1122**: 278-282.

Takase, K. (1993) Effect of mutation of an amino acid residue near the catalytic site on the activity of *Bacillus stearothermophilus* α-amylase. *Eur. J. Biochem.* **211**: 899-902.

Takizawa, N. and Murooka, Y. (1985) Cloning of the pullulanase gene and overproduction of pullulanase in *Escherichia coli* and *Klebsiella aerogenes*. *Appl. Environ. Microbiol.* **49**(2): 294-298.

Tarabasz-Szymanska, L., Galas, E. and Pankiewicz, T. (1999) Optimization of productivity of pullulan by means of multivariable linear regression analysis. *Enz. Microbial. Tech.* **24**: 276-282.

Thompson, C. R. (1998). α-amylase and pullulanase of *Thermopallium natronophilum*. *Department of Biology and Biochemistry*, University of Bath.

Thornton, J. M. and Taylor, W. R. *Structure prediction*, IRL Press: 147-190.

Tomiyasu, K., Yato, K., Yasuda, M., Tonozuka, T., Ibuka, A. and Sakai, H. (2001) Cloning and nucleotide sequence of the pullulanase gene of *Thermus thermophilus* HB8 and production of the enzyme in *Escherichia coli*. *Biosci. Biotech. Biochem.* **65**(9): 2090-2094.

Tonozuka, T., Ohtsuka, M., Mogi, S., Sakai, H., Ohta, T. and Sakano, Y. (1993) A neopullulanase-type alpha-amylase gene from *Thermoactinomyces vulgaris* R-47. *Biosci. Biotechnol. Biochem.* **57**(3): 395-401.

Uitdehaag, J. C. M., Veen, B. A., Dijkhuizen, L. and Dijkstra, B. W. (2002) Catalytic mechanism and product specificity of cyclodextrin glycosyltransferase, a prototypical transglycosylase from the α-amylase family. *Enz. Microbial. Tech.* **30**: 295-304.

Veen, B. A., Uitdehaag, J. C. M., Dijkstra, B. W. and Dijkhuizen, L. (2000) Engineering of cyclodextrin glycosyltransferase reaction and product specificity. *Biochim. Biophys. Acta* **1543**(2): 336-360.

Vihinen, M., Ollikka, P., Niskanen, J., Meyer, P., Suominen, I., Karp, M., Holm, L., Knowles, J. and Mantsala, P. (1990) Site-directed mutagenesis of a thermostable alpha-amylase from Bacillus stearothermophilus: putative role of three conserved residues. *J. Biochem.* **107**(2): 267-272.

Watabe, T., Miyake, K. and Suzuki, Y. (2001) Identification of catalytic and substrate-binding site residues in *Bacillus cereus* ATCC7064 oligo-1,6-glucosidase. *Biosci. Biotech. Biochem.* **65**(9): 2058-2064.

Watanabe, K., Kitamura, K., Iha, H. and Suzuki, Y. (1990) Primary structure of the oligo-1,6-glucosidase *Bacillus cereus* ATCC7064 deduced from the nucleotide sequence of the cloned gene. *Eur. J. Biochem.* **192**: 609-620.

Wiegal, J. (1998) Anaerobic alkalithermophiles, a novel group of extremophiles. *Extremophiles* **2**: 257-267.

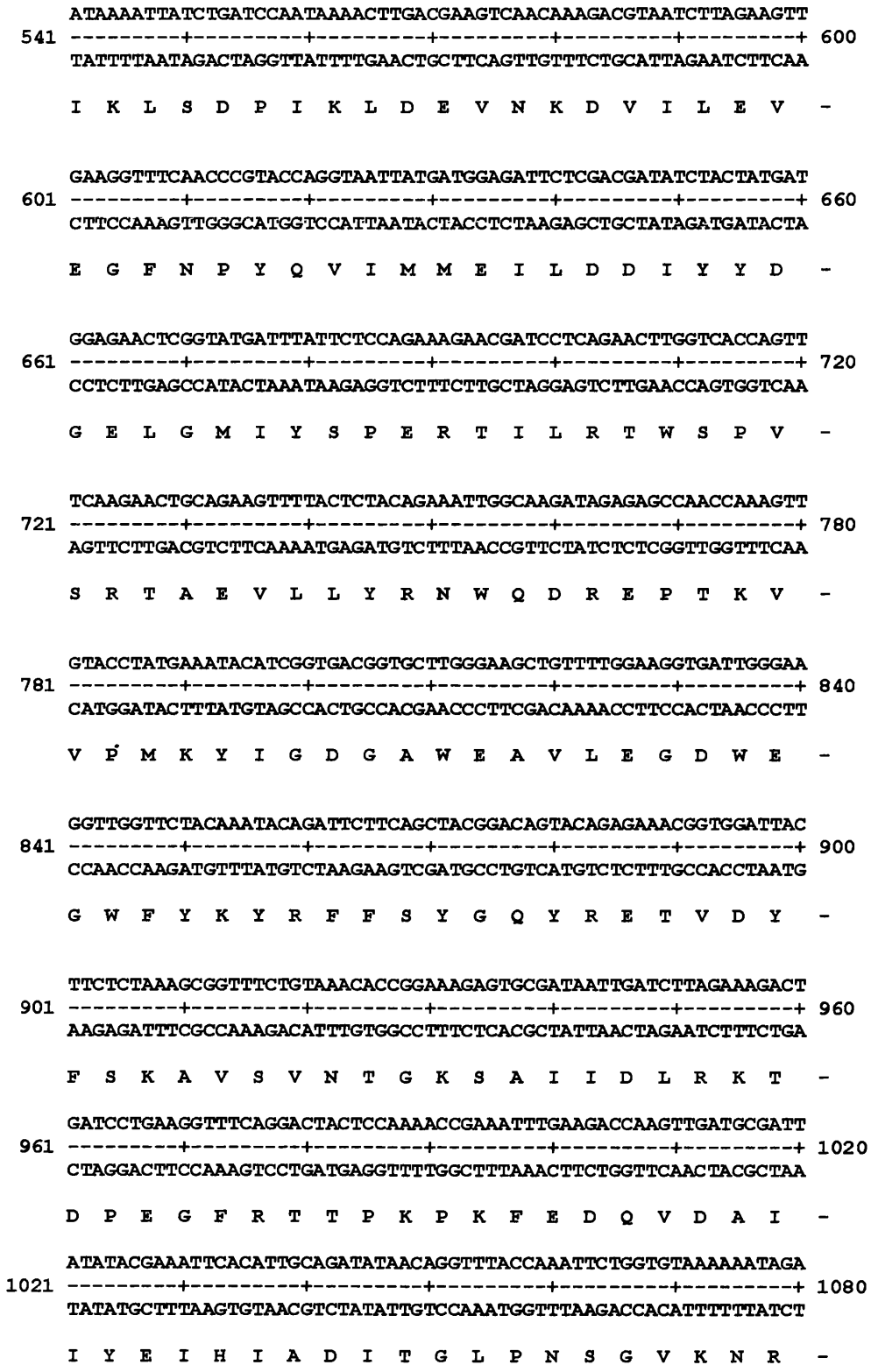
Yamashita, M., Kinoshita, T., Ihara, M., Mikawa, T. and Murooka, Y. (1994) Random mutagenesis of pullulanase from *Klebsiella aerogenes* for studies of the structure and function of the enzyme. *J. Biochem.* **116**: 1233-1240.

Yamashita, M., Matsumoto, D. and Murooka, Y. (1997) Amino acid residues specific for the catalytic action towards α-1,6-glucosidic linkages in *Klebsiella* pullulanase. *Journal of Fermentation and Bioengineering* **34**(4): 283-290.

Appendix I:

The Gene and Peptide Sequence of Pullulanase Type I from *T.natronophilum*

```
    ATGAAACGGATATTTTCGGTAGTTTTACTTTTAACGGTTGTATTTCTCTTTGCTCAAACT
  1 ---------+---------+---------+---------+---------+---------+ 60
    TACTTTGCCTATAAAAGCCATCAAAATGAAAATTGCCAACATAAAGAGAAACGAGTTTGA

    M  K  R  I  F  S  V  V  L  L  L  T  V  V  F  L  F  A  Q  T  -

    GAACTTATTATCCACTACCACAGATTTGATGGAAACTACGAAGGCTGGAACCTTTGGATA
 61 ---------+---------+---------+---------+---------+---------+ 120
    CTTGAATAATAGGTGATGGTGTCTAAACTACCTTTGATGCTTCCGACCTTGGAAACCTAT

    E  L  I  I  H  Y  H  R  F  D  G  N  Y  E  G  W  N  L  W  I  -

    TGGTACGTCGAACCGATATCAGCTGAAGGTAGAGCTTATCAGTTCACTGAAAAGGACGAT
121 ---------+---------+---------+---------+---------+---------+ 180
    ACCATGCAGCTTGGCTATAGTCGACTTCCATCTCGAATAGTCAAGTGACTTTTCCTGCTA

    W  Y  V  E  P  I  S  A  E  G  R  A  Y  Q  F  T  E  K  D  D  -

    TTTGGTGTTGTTGCAAGAATAGTTTTACCAGAGACGTTGACAAGAGTTGGTATTATTGTT
181 ---------+---------+---------+---------+---------+---------+ 240
    AAACCACAACAACGTTCTTATCAAAATGGTCTCTGCAACTGTTCTCAACCATAATAACAA

    F  G  V  V  A  R  I  V  L  P  E  T  L  T  R  V  G  I  I  V  -

    AGACTTGGTGAATGGGTAATGAAAGATGTCGCAATGGACAGGTTCATAGACATTAAAGAT
241 ---------+---------+---------+---------+---------+---------+ 300
    TCTGAACCACTTACCCATTACTTTCTACAGCGTTACCTGTCCAAGTATCTGTAATTTCTA

    R  L  G  E  W  V  M  K  D  V  A  M  D  R  F  I  D  I  K  D  -

    GGAAAAGCGGAGGTGTGGCTCTTGCAAGGAGTTGAACAAATATATACAAGCAGACCAGAT
301 ---------+---------+---------+---------+---------+---------+ 360
    CCTTTTCGCCTCCACACCGAGAACGTTCCTCAACTTGTTTATATATGTTCGTCTGGTCTA

    G  K  A  E  V  W  L  L  Q  G  V  E  Q  I  Y  T  S  R  P  D  -

    ACAAACCCAAGAGTATTCTTTGCACAAGCGAGAAAAACTGATGTAATTGAAGCTTATTTA
361 ---------+---------+---------+---------+---------+---------+ 420
    TGTTTGGGTTCTCATAAGAAACGTGTTCGCTCTTTTTGACTACATTAACTTCGAATAAAT

    T  N  P  R  V  F  F  A  Q  A  R  K  T  D  V  I  E  A  Y  L  -

    ACACATCCTGTTGATACGAGAACAGCTGAAGTAAAAGTTACGGTAGACGGAGAAGAAAGA
421 ---------+---------+---------+---------+---------+---------+ 480
    TGTGTAGGACAACTATGCTCTTGTCGACTTCATTTTCAATGCCATCTGCCTCTTCTTTCT

    T  H  P  V  D  T  R  T  A  E  V  K  V  T  V  D  G  E  E  R  -

    AAGATAAAAATCTGTAGAAAGAGCTGATCCAACGGATATTTCAGTTACAAATTGGATTAGG
481 ---------+---------+---------+---------+---------+---------+ 540
    TTCTATTTTAGACATCTTTCTCGACTAGGTTGCCTATAAAGTCAATGTTTAACCTAATCC

    K  I  K  S  V  E  R  A  D  P  T  D  I  S  V  T  N  W  I  R  -
```


ATAAAATTATCTGATCCAATAAAACTTGACGAAGTCAACAAAGACGTAATCTTAGAAGTT
541 ---------+---------+---------+---------+---------+---------+ 600
TATTTTAATAGACTAGGTTATTTTGAACTGCTTCAGTTGTTTCTGCATTAGAATCTTCAA
I K L S D P I K L D E V N K D V I L E V -
GAAGGTTTCAACCCGTACCAGGTAATTATGATGGAGATTCTCGACGATATCTACTATGAT
601 ---------+---------+---------+---------+---------+---------+ 660
CTTCCAAAGTTGGGCATGGTCCATTAATACTACCTCTAAGAGCTGCTATAGATGATACTA
E G F N P Y Q V I M M E I L D D I Y Y D -
GGAGAACTCGGTATGATTTATTCTCCAGAAAGAACGATCCTCAGAACTTGGTCACCAGTT
661 ---------+---------+---------+---------+---------+---------+ 720
CCTCTTGAGCCATACTAAATAAGAGGTCTTTCTTGCTAGGAGTCTTGAACCAGTGGTCAA
G E L G M I Y S P E R T I L R T W S P V -
TCAAGAACTGCAGAAGTTTTACTCTACAGAAATTGGCAAGATAGAGAGCCAACCAAAGTT
721 ---------+---------+---------+---------+---------+---------+ 780
AGTTCTTGACGTCTTCAAAATGAGATGTCTTTAACCGTTCTATCTCTCGGTTGGTTTCAA
S R T A E V L L Y R N W Q D R E P T K V -
GTACCTATGAAATACATCGGTGACGGTGCTTGGGAAGCTGTTTTGGAAGGTGATTGGGAA
781 ---------+---------+---------+---------+---------+---------+ 840
CATGGATACTTTATGTAGCCACTGCCACGAACCCTTCGACAAAACCTTCCACTAACCCTT
V P M K Y I G D G A W E A V L E G D W E -
GGTTGGTTCTACAAATACAGATTCTTCAGCTACGGACAGTACAGAGAAACGGTGGATTAC
841 ---------+---------+---------+---------+---------+---------+ 900
CCAACCAAGATGTTTATGTCTAAGAAGTCGATGCCTGTCATGTCTCTTTGCCACCTAATG
G W F Y K Y R F F S Y G Q Y R E T V D Y -
TTCTCTAAAGCGGTTTCTGTAAACACCGGAAAGAGTGCGATAATTGATCTTAGAAAGACT
901 ---------+---------+---------+---------+---------+---------+ 960
AAGAGATTTCGCCAAAGACATTTGTGGCCTTTCTCACGCTATTAACTAGAATCTTTCTGA
F S K A V S V N T G K S A I I D L R K T -
GATCCTGAAGGTTTCAGGACTACTCCAAAACCGAAATTTGAAGACCAAGTTGATGCGATT
961 ---------+---------+---------+---------+---------+---------+ 1020
CTAGGACTTCCAAAGTCCTGATGAGGTTTTGGCTTTAAACTTCTGGTTCAACTACGCTAA
D P E G F R T T P K P K F E D Q V D A I -
ATATACGAAATTCACATTGCAGATATAACAGGTTTACCAAATTCTGGTGTAAAAAATAGA
1021 ---------+---------+---------+---------+---------+---------+ 1080
TATATGCTTTAAGTGTAACGTCTATATTGTCCAAATGGTTTAAGACCACATTTTTTATCT
I Y E I H I A D I T G L P N S G V K N R -

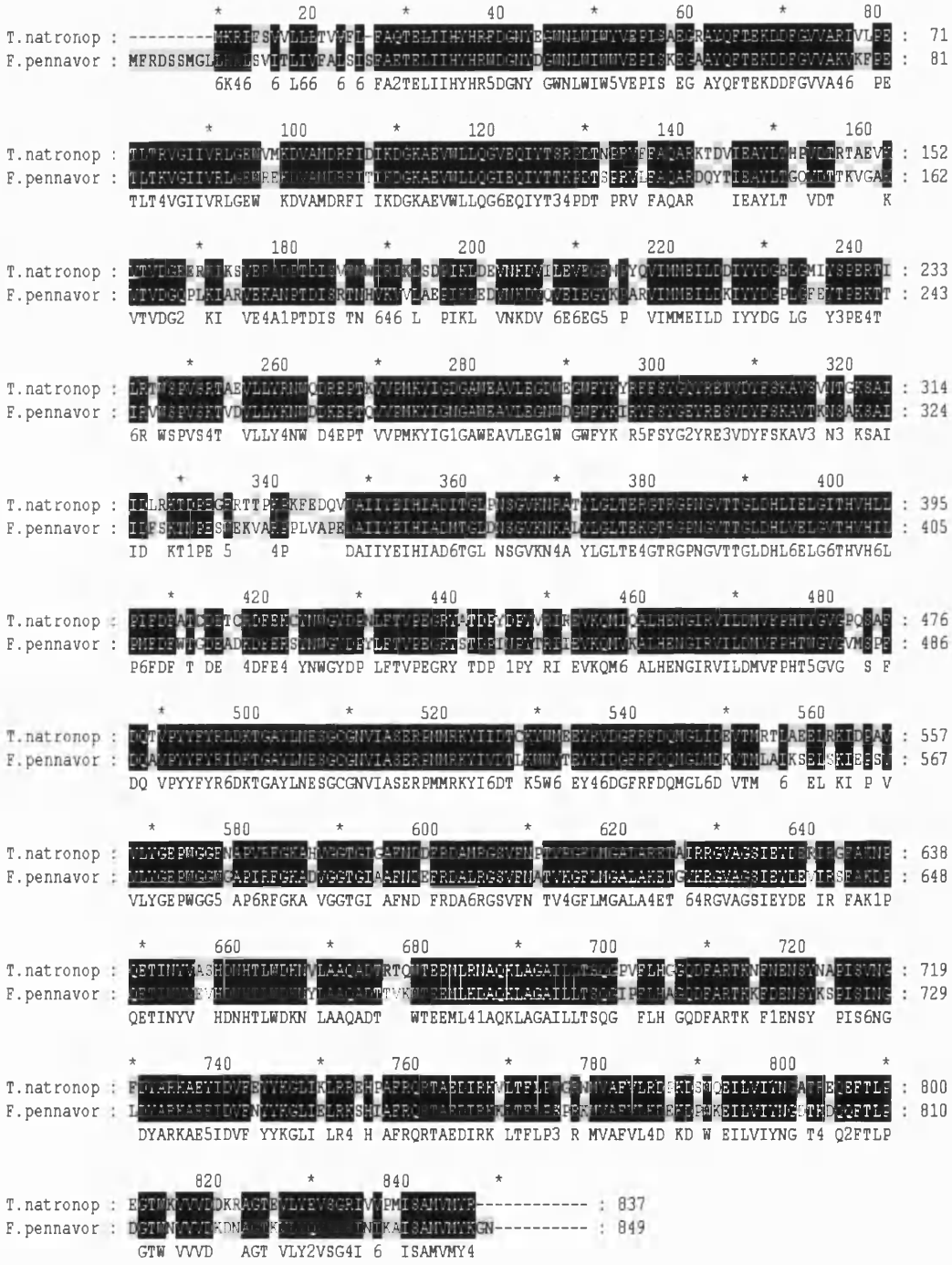
```
     GCAACTTATTTAGGATTAACCGAAAGGGGTACAAGAGGACCGAATGGAGTCACAACCGGG
1081 ---------+---------+---------+---------+---------+---------+ 1140
     CGTTGAATAAATCCTAATTGGCTTTCCCCATGTTCTCCTGGCTTACCTCAGTGTTGGCCC

     A  T  Y  L  G  L  T  E  R  G  T  R  G  P  N  G  V  T  T  G  -

     CTTGACCATCTAATCGAACTTGGAATTACACACGTTCACTTACTTCCTATATTTGACTTT
1141 ---------+---------+---------+---------+---------+---------+ 1200
     GAACTGGTAGATTAGCTTGAACCTTAATGTGTGCAAGTGAATGAAGGATATAAACTGAAA

     L  D  H  L  I  E  L  G  I  T  H  V  H  L  L  P  I  F  D  F  -

     GCAACTTGTGACGAAACTTGTAGAGATTTTGAAAAATGCTACAACTGGGGCTATGATCCA
1201 ---------+---------+---------+---------+---------+---------+ 1260
     CGTTGAACACTGCTTTGAACATCTCTAAAACTTTTTACGATGTTGACCCCGATACTAGGT

     A  T  C  D  E  T  C  R  D  F  E  K  C  Y  N  W  G  Y  D  P  -

     AATCTTTTTACCGTTCCAGAAGGTAGATACGCAACAGATCCGTATGATCCTTATGTAAGA
1261 ---------+---------+---------+---------+---------+---------+ 1320
     TTAGAAAAATGGCAAGGTCTTCCATCTATGCGTTGTCTAGGCATACTAGGAATACATTCT

     N  L  F  T  V  P  E  G  R  Y  A  T  D  P  Y  D  P  Y  V  R  -

     ATAAGAGAAGTTAAGCAAATGATTCAAGCACTCCACGAAAATGGAATAAGAGTCATTCTC
1321 ---------+---------+---------+---------+---------+---------+ 1380
     TATTCTCTTCAATTCGTTTACTAAGTTCGTGAGGTGCTTTTACCTTATTCTCAGTAAGAG

     I  R  E  V  K  Q  M  I  Q  A  L  H  E  N  G  I  R  V  I  L  -

     GATATGGTCTTTCCACACACATATGGTGTAGGTCCTCAATCAGCTTTTGACCAAACAGTT
1381 ---------+---------+---------+---------+---------+---------+ 1440
     CTATACCAGAAAGGTGTGTGTATACCACATCCAGGAGTTAGTCGAAAACTGGTTTGTCAA

     D  M  V  F  P  H  T  Y  G  V  G  P  Q  S  A  F  D  Q  T  V  -

     CCATATTATTTCTACAGACTTGACAAAACAGGAGCTTATCTCAACGAAAGTGGTTGTGGA
1441 ---------+---------+---------+---------+---------+---------+ 1500
     GGTATAATAAAGATGTCTGAACTGTTTTGTCCTCGAATAGAGTTGCTTTCACCAACACCT

     P  Y  Y  F  Y  R  L  D  K  T  G  A  Y  L  N  E  S  G  C  G  -

     AATGTTATCGCTTCCGAAAGACCGATGATGAGGAAATATATAATTGATACTTGTAAATAC
1501 ---------+---------+---------+---------+---------+---------+ 1560
     TTACAATAGCGAAGGCTTTCTGGCTACTACTCCTTTATATATTAACTATGAACATTTATG

     N  V  I  A  S  E  R  P  M  M  R  K  Y  I  I  D  T  C  K  Y  -

     TGGATGGAAGAGTACAGAGTAGACGGTTTCAGATTTGACCAAATGGGACTCATTGACGAA
1561 ---------+---------+---------+---------+---------+---------+ 1620
     ACCTACCTTCTCATGTCTCATCTGCCAAAGTCTAAACTGGTTTACCCTGAGTAACTGCTT

     W  M  E  E  Y  R  V  D  G  F  R  F  D  Q  M  G  L  I  D  E  -
```

```
     GTCACAATGAGAACACTTGCAGAAGAACTCAGAAAGATTGACCCTGCGGTAGTTCTCTAC
1621 ---------+---------+---------+---------+---------+---------+ 1680
     CAGTGTTACTCTTGTGAACGTCTTCTTGAGTCTTTCTAACTGGGACGCCATCAAGAGATG

     V  T  M  R  T  L  A  E  E  L  R  K  I  D  P  A  V  V  L  Y  -


     GGAGAACCTTGGGGCGGTTTCAACGCACCTGTCAGATTCGGGAAAGCACACGTCGGTGGC
1681 ---------+---------+---------+---------+---------+---------+ 1740
     CCTCTTGGAACCCCGCCAAAGTTGCGTGGACAGTCTAAGCCCTTTCGTGTGCAGCCACCG

     G  E  P  W  G  G  F  N  A  P  V  R  F  G  K  A  H  V  G  G  -


     ACTGGAATCGGAGCTTTCAACGACGATTTCAGAGACGCTATGAGAGGTTCTGTCTTTAAT
1741 ---------+---------+---------+---------+---------+---------+ 1800
     TGACCTTAGCCTCGAAAGTTGCTGCTAAAGTCTCTGCGATACTCTCCAAGACAGAAATTA

     T  G  I  G  A  F  N  D  D  F  R  D  A  M  R  G  S  V  F  N  -


     CCAACAGTTAGAGGTTTCTTGATGGGAGCACTTGCAAGAGAAACAGCGATAAGAAGAGGA
1801 ---------+---------+---------+---------+---------+---------+ 1860
     GGTTGTCAATCTCCAAAGAACTACCCTCGTGAACGTTCTCTTTGTCGCTATTCTTCTCCT

     P  T  V  R  G  F  L  M  G  A  L  A  R  E  T  A  I  R  R  G  -


     GTTGCAGGAAGTATTGAATACGACGAACGAATTAGAGGCTTTGCTAAGAACCCACAGGAA
1861 ---------+---------+---------+---------+---------+---------+ 1920
     CAACGTCCTTCATAACTTATGCTGCTTGCTTAATCTCCGAAACGATTCTTGGGTGTCCTT

     V  A  G  S  I  E  Y  D  E  R  I  R  G  F  A  K  N  P  Q  E  -


     ACGATAAATTATGTTGCATCACACGATAATCACACACTTTGGGACAAAAACGTATTAGCA
1921 ---------+---------+---------+---------+---------+---------+ 1980
     TGCTATTTAATACAACGTAGTGTGCTATTAGTGTGTGAAACCCTGTTTTTGCATAATCGT

     T  I  N  Y  V  A  S  H  D  N  H  T  L  W  D  K  N  V  L  A  -


     GCTCAAGCGGATACGAGAACTCAATGGACAGAGGAAATGCTTAGAAACGCTCAAAAACTT
1981 ---------+---------+---------+---------+---------+---------+ 2040
     CGAGTTCGCCTATGCTCTTGAGTTACCTGTCTCCTTTACGAATCTTTGCGAGTTTTTGAA

     A  Q  A  D  T  R  T  Q  W  T  E  E  M  L  R  N  A  Q  K  L  -


     GCGGGCGCAATACTTCTAACATCTCAAGGACCAGTCTTTTTGCACGGTGGTCAAGACTTT
2041 ---------+---------+---------+---------+---------+---------+ 2100
     CGCCCGCGTTATGAAGATTGTAGAGTTCCTGGTCAGAAAAACGTGCCACCAGTTCTGAAA

     A  G  A  I  L  L  T  S  Q  G  P  V  F  L  H  G  G  Q  D  F  -


     GCAAGAACGAAGAATTTTAATGAAAACTCGTACAACGCACCTATTTCTGTTAACGGTTTT
2101 ---------+---------+---------+---------+---------+---------+ 2160
     CGTTCTTGCTTCTTAAAATTACTTTTGAGCATGTTGCGTGGATAAAGACAATTGCCAAAA

     A  R  T  K  N  F  N  E  N  S  Y  N  A  P  I  S  V  N  G  F  -
```

```
     GACTACGCAAGAAAAGCTGAATATATCGATGTGTTTGAATACTACAAAGGTCTTATCAAG
2161 ---------+---------+---------+---------+---------+---------+ 2220
     CTGATGCGTTCTTTTCGACTTATATAGCTACACAAACTTATGATGTTTCCAGAATAGTTC

     D  Y  A  R  K  A  E  Y  I  D  V  F  E  Y  Y  K  G  L  I  K   -


     CTTAGAAGAGAACATCCAGCGTTCAGACAAAGAACTGCGGAAGATATTAGGAAAGTTTTA
2221 ---------+---------+---------+---------+---------+---------+ 2280
     GAATCTTCTCTTGTAGGTCGCAAGTCTGTTTCTTGACGCCTTCTATAATCCTTTCAAAAT

     L  R  R  E  H  P  A  F  R  Q  R  T  A  E  D  I  R  K  V  L   -


     ACCTTCTTACCAACAGGAAGAAATATGGTTGCATTCGTCTTGAGAGATCCGAAAGATAGT
2281 ---------+---------+---------+---------+---------+---------+ 2340
     TGGAAGAATGGTTGTCCTTCTTTATACCAACGTAAGCAGAACTCTCTAGGCTTTCTATCA

     T  F  L  P  T  G  R  N  M  V  A  F  V  L  R  D  P  K  D  S   -


     TGGCAAGAAATTCTTGTGATCTACAATGGGGCAACAAGAGAACAAGAATTTACACTTCCA
2341 ---------+---------+---------+---------+---------+---------+ 2400
     ACCGTTCTTTAAGAACACTAGATGTTACCCCGTTGTTCTCTTGTTCTTAAATGTGAAGGT

     W  Q  E  I  L  V  I  Y  N  G  A  T  R  E  Q  E  F  T  L  P   -


     GAAGGTACTTGGAAAGTTGTAGTTGATGATAAAAGGGCAGGTACTGAGGTCCTTTACGAA
2401 ---------+---------+---------+---------+---------+---------+ 2460
     CTTCCATGAACCTTTCAACATCAACTACTATTTTCCCGTCCATGACTCCAGGAAATGCTT

     E  G  T  W  K  V  V  V  D  D  K  R  A  G  T  E  V  L  Y  E   -


     GTGAGCGGTAGAATTGTAGTACCAATGATTAGCGCAATGGTGATGTACAGATAA
2461 ---------+---------+---------+---------+---------+---- 2514
     CACTCGCCATCTTAACATCATGGTTACTAATCGCGTTACCACTACATGTCTATT

     V  S  G  R  I  V  V  P  M  I  S  A  M  V  M  Y  R  *   -
```

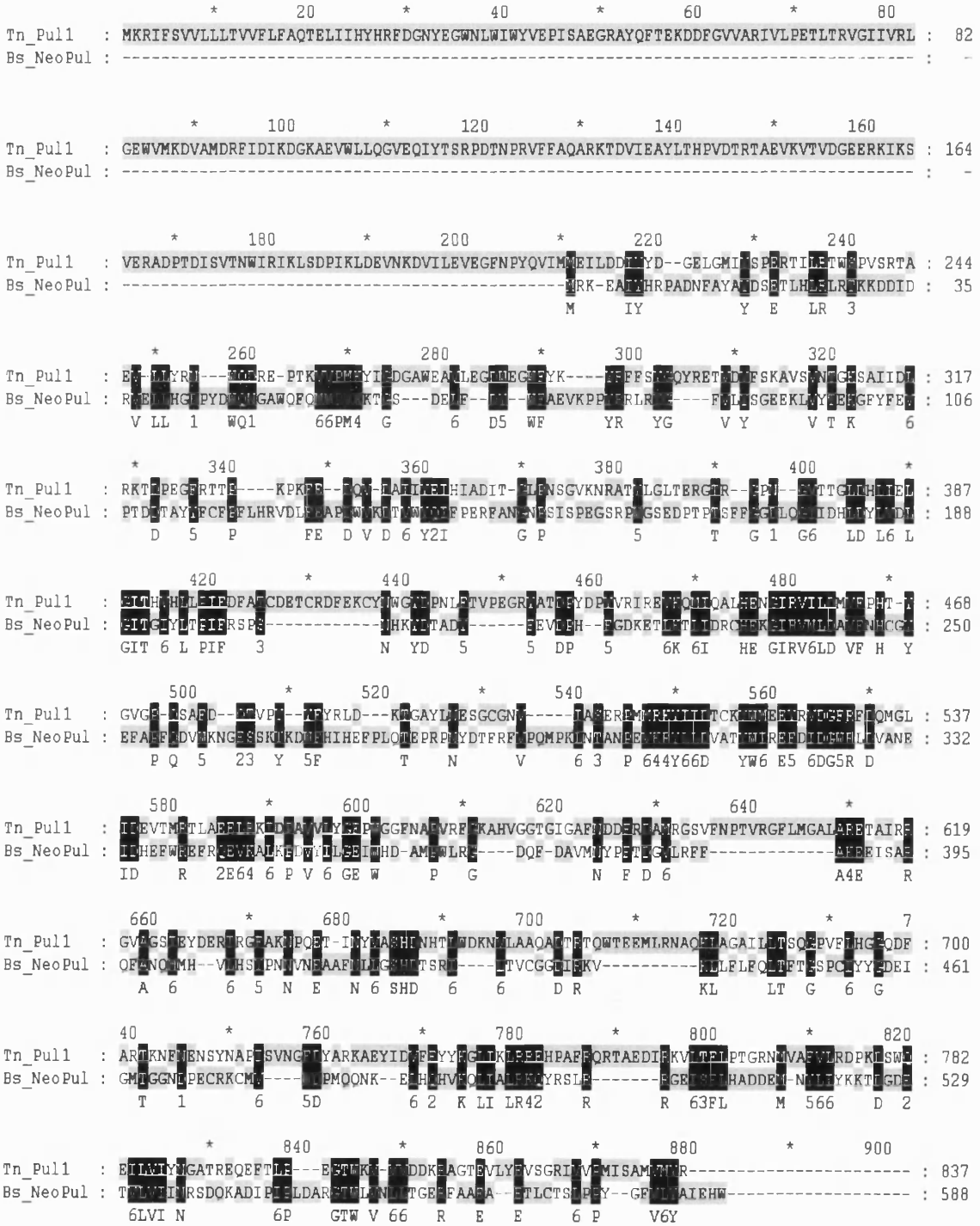
Appendix II

Sequence Alignment of Pullulanase Type I from *T.natronophilum* and *F.pennavorans* Ven5



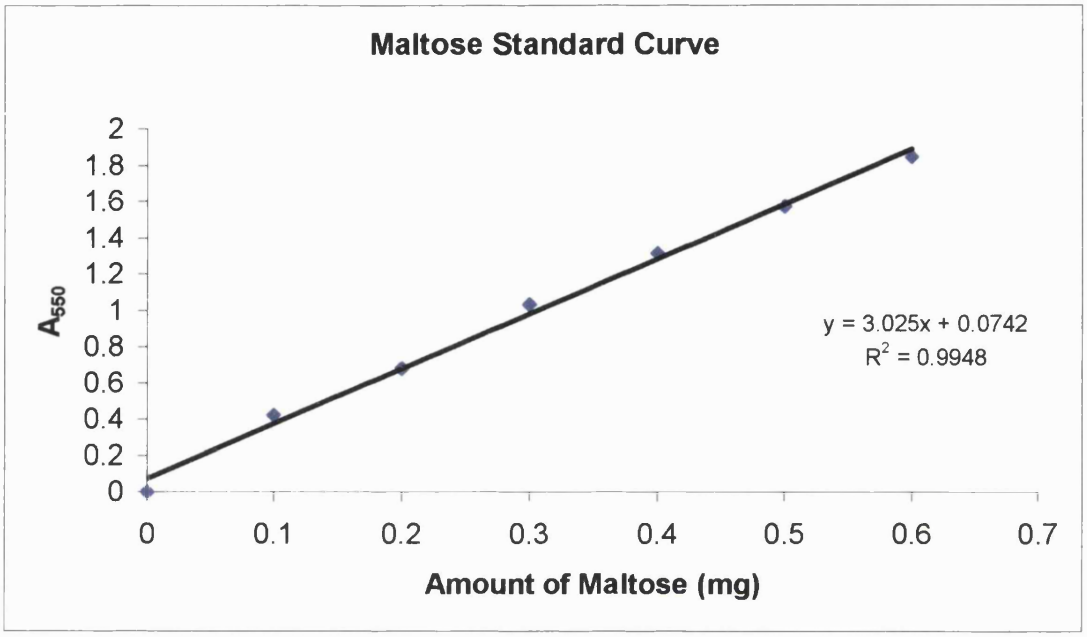
Appendix III:

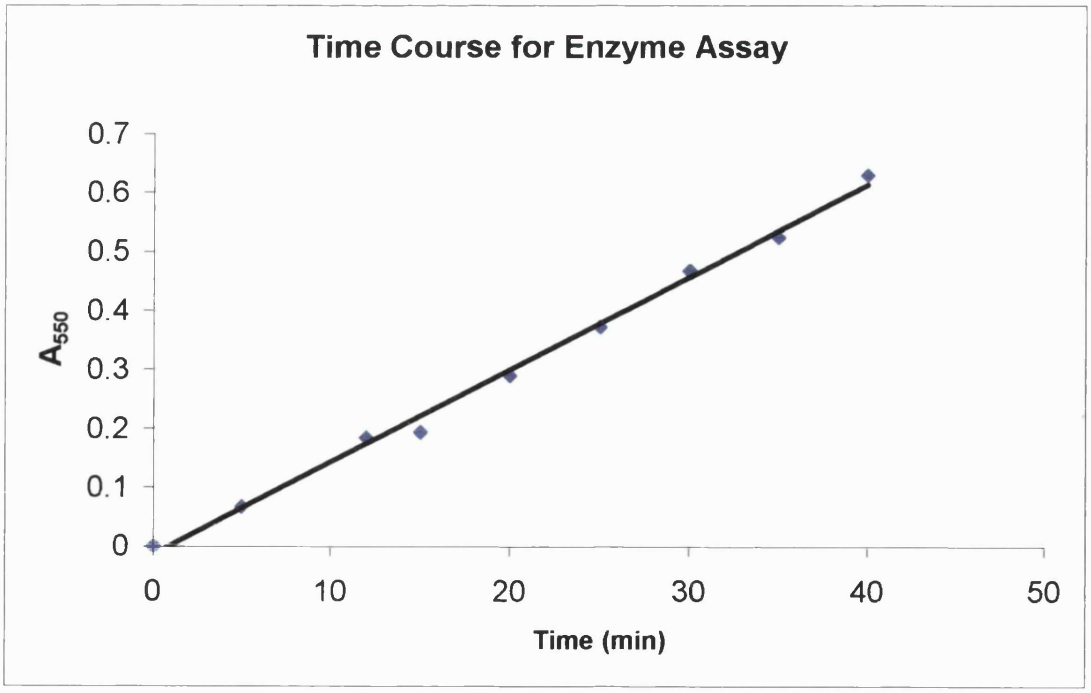
Sequence Alignment of *B.stearothermophilus* Neopullulanase with *T.natronophilum* Pullulanase

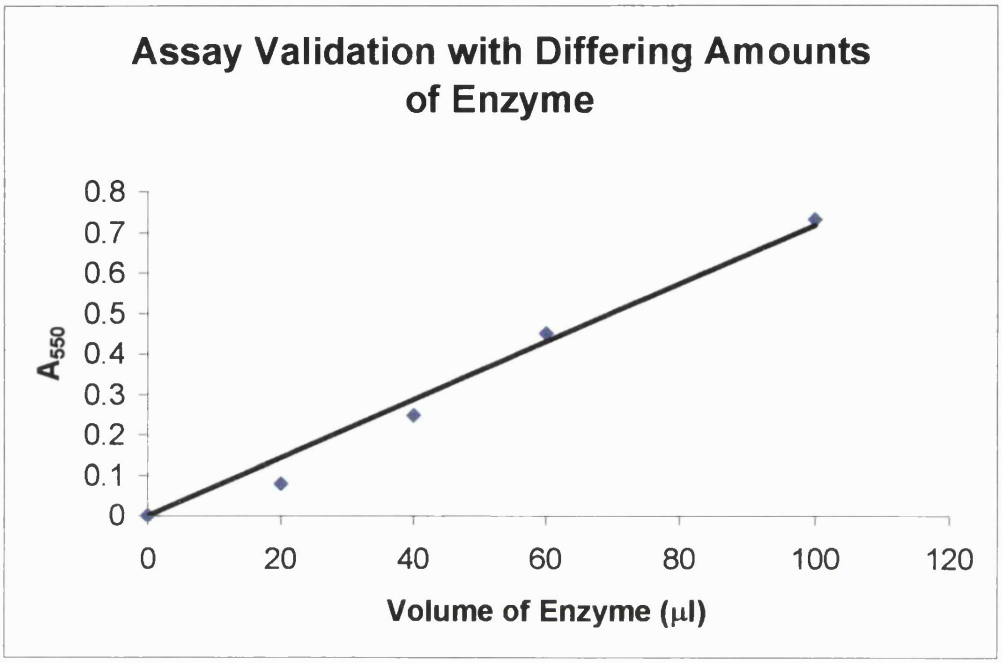


Appendix IV

Assay Validation






Assay Validation with Differing Amounts of Enzyme
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
A550
0
20
40
60
80
100
120
Volume of Enzyme (μl)